

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 3, 2004, 04:57:13 ; Search time 707.789 seconds
(without alignments)
2110.938 Million cell updates/sec

Title: US-09-933-267A-1_COPY_306362_306402

Perfect score: 41

Sequence: 1 ttgttgatgctgagccccc.....atactctatccgagtatga 41

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*

1: gb_est1:*
2: gb_est2:*
3: gb_hic:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gss1:*
9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	41	100.0	611	5	BX489897	BX489897 DKFZp686N
2	39.4	96.1	653	1	AI073549	AI073549 ov45e08.x
3	37.8	92.2	705	5	BP160291	BP160291 BP160291
4	31.4	76.6	593	1	AA023625	AA023625 mh80b06.r
5	31.4	76.6	616	4	BG084311	BG084311 H3099A08-
6	31.4	76.6	627	2	BH200714	BH200714 B3200714
7	31.4	76.6	2109	3	AK039911	AK039911 Mus muscu
8	31.4	76.6	2591	3	AK077236	AK077236 Mus muscu
9	31.4	76.6	2735	3	AK054182	AK054182 Mus muscu
10	31.4	76.6	4248	3	AK041525	AK041525 Mus muscu
11	31.4	76.6	4315	3	AK087638	AK087638 Mus muscu
12	31.4	76.6	4321	3	AK036627	AK036627 Mus muscu
13	29	70.7	365	1	AA092184	AA092184 114609.se
14	28.8	70.2	576	1	AU041270	AU041270 AU041270
15	27.8	67.8	656	9	CG483529	CG483529 OST17199
16	23.4	57.1	1169	5	BU588298	BU588298 AGNCOURT
17	23	56.1	871	8	B2172878	B2172878 CH230-505
18	22.6	55.1	748	9	AG356459	AG356459 Mus muscu
19	22.4	54.6	213	2	AW582379	AW582379 MR3-ST022
20	22.4	54.6	599	8	AZ833415	AZ833415 2M0115A06
21	22.4	54.6	646	9	CE331215	CE331215 tigr-gss-
22	22	53.7	481	6	CA993426	CA993426 PAR07E12
23	22	53.7	563	6	CA917333	CA917333 EST641480
24	22	53.7	904	9	CG937676	CG937676 MBEMW13TF

CG611862 OST297347
BU874710 Q071A12 P
CE849311 tigr-gss-
CO023851 EST788984
CA116396 SCACLR105
CA100736 SCCCL700
CA261912 SCBGLB201
BG145973 uu95B10.Y
BF922664 QV4-NT025
CO272045 EK105718.
BQ031691 UI-1-CF0-
CD291168 StrPu538.
CK134778 RH01453.3
BU957489 AGENCOURT
AK047386 Mus muscu
AQ985336 RPCI-23-3
CD727728 4032587 1
CK981913 4114536 B
AZ858611 2M0163D19
BM909339 505541 MA
CK941792 4065309 B

ALIGNMENTS

RESULT 1
BX489897

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BX489897 611 bp mRNA linear EST 04-SEP-2003
DKFZp686N0562 r1 686 (synonym: hlccc3) Homo sapiens cDNA clone
DKFZp686N0562 5', mRNA sequence.

ACCESSION BX489897.1 GI:31998955

VERSION EST.

KEYWORDS Homo sapiens (human)

SOURCE Homo sapiens

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CONTACT: MIPs

MIPs

Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany

This is the 5' sequence of the clone insert

Clone from S. Wiemann, Molecular Genome Analysis, German Cancer

Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;

sequenced by AGOWA (Berlin/Germany) within the cDNA sequencing

consortium of the German Genome Project.

No sl sequence available.

This clone (DKFZp686N0562) is available at the RZPD in Berlin.

Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059

Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

Location/Qualifiers

1..611

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="DKFZp686N0562"

/dev_stage="adult"

/lab_host="Dros"

/clone_lib="686 (synonym: hlccc3)"

/notes="Vector: pTriplex2; Site_1: SfiIA; Site_2: SfiIB;

cDNA-collection"

ORIGIN

Query Match

Best Local Similarity

Matches

100.0%; Score 41; DB 5; Length 611;

100.0%; Pred. No. 7.8e-06;

Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGTTGATGCTGAGCCCCCACTACTATTCGAGTATGA 41

```

|||||
199 TTGTTGGATGCTGAGCCCCCATCTACTTATCCGAGTATGA 239

RESULT 2
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
1 (bases 1 to 705)
Unishu,H., Eguchi,T., Suzuki,K., Sawazaki,T., Toki,D., Shinkai,H.,
Okumura,N., Hamasima,N. and Awata,T.
PEDE (Pig EST Data Explorer): construction of a database for ESTs
derived from porcine full-length cDNA libraries
Nucleic Acids Res. 32 (1), D484-D488 (2004)
Contact: Hirohide Venishi
Animal Genome Laboratory, Genome Research Department
National Institute of Agrobiological Sciences
2 Ikenodai, Tsukuba, Ibaraki 305-8602, Japan
Tel: +81-29-838-8627
Fax: +81-29-838-8627
Email: huenishiaffair.go.jp
EST project with full-length enriched cDNA libraries carried out in
Animal Genome Research Program (Japan) by National Institute of
Agrobiological Sciences and STAFF-Institute
Single pass sequencing of clones derived from oligo-capped cDNA
library
Vector sequences were eliminated by RepeatMasker version 2002/07/13
and crossmatch version 0.990319
Low quality bases were trimmed based on the quality values.

FEATURES
source
1..705
/organism="Sus scrofa"
/mol_type="mRNA"
/db_xref="taxon:9823"
/clone="THY010029A10"
/tissue_type="thymus"
/dev_stage="adult"
/clone_lib="full-length enriched swine cDNA library, adult
thymus"

ORIGIN
Query Match 92.2%; Score 37.8; DB 5; Length 705;
Best Local Similarity 95.1%; Pred. No. 0.00016;
Matches 39; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TTGTTGGATGCTGAGCCCCCATCTACTTATCCGAGTATGA 41
|||||
Db 553 TTGTTGGAGCTGAGCCCCCATCTACTTATCCGAGTATGA 593

RESULT 4
LOCUS
DEFINITION
AA023625/c 593 bp mRNA linear EST 21-JAN-1997
mh80b06.r1 Soares mouse placenta 4NMPI3.5 14.5 Mus musculus cDNA
clone IMAGE:457235 5' similar to gb:M12674 ESTROGEN RECEPTOR
(HUMAN); gb:M38651 Mouse estrogen receptor mRNA, complete cds
(MOUSE); mRNA sequence.
ACCESSION
VERSION
AA023625.1 GI:1487542
KEYWORDS
SOURCE
Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 593)
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wyllie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
The WashU-HMI Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu

TITLE
JOURNAL
COMMENT

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 653)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: -40mi3 fwd. ET from Amersham
High quality sequence stop: 483.

FEATURES
source
1..653
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE.1640294"
/sex="male"
/lab_host="DH10B"
/clone_lib="Soares testis_NHT"
/notes="Vector: pT73D-pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech
Laboratories, Inc., and primed with a Not I - oligo(dT)
primer [5'
TGTTACCAATCTGAGTGGAGCGCGCCCACTTTTCTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization to Cot5, and was
constructed by Bento Soares and M. Fatima Bonaldo."

ORIGIN
Query Match 96.1%; Score 39.4; DB 1; Length 653;
Best Local Similarity 97.6%; Pred. No. 3.5e-05;
Matches 40; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TTGTTGGATGCTGAGCCCCCATCTACTTATCCGAGTATGA 41
|||||
Db 458 TTGTTGGATGCTGAGCCCCGATCTACTTATCCGAGTATGA 418

RESULT 3
BP160291 705 bp mRNA linear EST 30-DEC-2003
LOCUS
DEFINITION
BP160291 full-length enriched swine cDNA library, adult thymus Sus
scrofa cDNA clone THY010029A10 5', mRNA sequence.
ACCESSION
VERSION
BP160291.1 GI:40409764
KEYWORDS
SOURCE
Sus scrofa (pig)
ORGANISM
Sus scrofa

```

This clone is available royalty-free through LML ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

MG1:274123

Seq primer: -28M13 rev2 from Amerham

High quality sequence stop: 492.

FEATURES

Location/Qualifiers

source

1..593

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="IMAGE:457235"

/sex="unknown"

/tissue_type="placenta"

/dev_stage="adult"

/lab_host="DH10B"

/clone_lib="Soares mouse placenta 4NDMP13.5 14.5"

/note="organ: placenta; Vector: p773D-Pac (Pharmacia)

with a modified polylinker; Site_1: Not I; Site 2: Eco RI;

1st strand cDNA was primed with a Not I - oligo(dT) primer

15'

TGTTACCACTGCTGAAGTGGAGCGCGCCGGAATTTTTTTTTTTTTTTTTT

T 3'; double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia), digested with Not I and cloned into the Not I

and Eco RI sites of the modified p773 vector. Library

went through one round of normalization, and was

constructed by Bento Soares and M.Fatima Bonaldo."

ORIGIN

Query Match

Best Local Similarity 76.6%; Score 31.4; DB 1; Length 593;

Matches 35; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy

1 TTGTTGGATGCTGAGCGCCCGCATCTCTATTCCGAGTATGA 41

Db

306 TTGTTGGATGCTGAGCGCCCGCATCTCTATTCCGAGTATGA 266

RESULT 5

BG084311

LOCUS

DEFINITION H3099A08-5 NIA Mouse 15K cDNA Clone Set Mus musculus cDNA clone

616 bp mRNA linear EST 18-DEC-2003

H3099A08 5', mRNA sequence.

ACCESSION

BG084311

VERSION

BG084311.2

KEYWORDS

EST.

SOURCE

Mus musculus

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

Tanaka,T.S., Jaradat,S.A., Lim,M.K., Kargul,G.J., Wang,X.,

Grahovac,M.J., Pantano,S., Sano,Y., Piao,Y., Nagaraja,R., Doi,H.,

Wood,W.H. III, Becker,X.G. and Ko,M.S.H.

Genome-wide expression profiling of mid-gestation placenta and

embryo using a 15,000 mouse developmental cDNA microarray

Proc. Natl. Acad. Sci. U.S.A. 97 (16), 9127-9132 (2000)

20381348

PURVED

10922068

COMMENT

On Jan 26, 2001 this sequence version replaced gi:12566875.

Other_ESTs: H3099A08-3

Contact: George J. Kargul

Laboratory of Genetics

National Institute on Aging/National Institutes of Health

333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA

Email: cdna@nslu.gsc.nia.nih.gov

This clone set has been freely distributed to the community. Please

visit <http://lgsun.gsc.nia.nih.gov/cdna/15k.html> for details.

Plate: H3099 row: A column: 08

Seq primer: -21M13 Reverse

High quality sequence stop: 616

POLYA-No.

Location/Qualifiers

source

1..616

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6J"

/db_xref="niaEST:H3099A08-5"

/db_xref="taxon:10090"

/clone="H3099A08"

/sex="Clones arrayed from a variety of cDNA libraries"

/dev_stage="Clones arrayed from a variety of cDNA libraries"

/lab_host="DH10B"

/clone_lib="NIA Mouse 15K cDNA Clone Set"

/note="Vector: pSPORT1; Site_1: SalI; Site 2: NotI; This

clone is among a rearranged set of 15,247 clones from 11

embryo cDNA libraries (including preimplantation stage

embryos from unfertilized egg to blastocyst, embryonic

part of E7.5 embryos, extraembryonic part of E7.5

embryos, and E12.5 female mesonephros/gonad) and one

newborn ovary cDNA library. Average insert size 1.5 kb.

All source libraries are cloned unidirectionally with

Oligo(dT)-Not primers. References include: (1)

Genome-wide expression profiling of mid-gestation

placenta and embryo using a 15,000 mouse developmental

cDNA microarray, 2000, Proc. Natl. Acad. Sci. U.S.A., 97:

9127-9132; (2) Large-scale cDNA analysis reveals phased

gene expression patterns during preimplantation mouse

development, 2000, Development, 127: 1737-1749; (3)

Genome-wide mapping of unselected transcripts from

extraembryonic tissue of 7.5-day mouse embryos reveals

enrichment in the t-complex and under-representation on

the X chromosome, 1998, Hum Mol Genet 7: 1967-1978."

ORIGIN

Query Match

Best Local Similarity 76.6%; Score 31.4; DB 4; Length 616;

Matches 35; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy

1 TTGTTGGATGCTGAGCGCCCGCATCTCTATTCCGAGTATGA 41

Db

371 TTGTTGGATGCTGAGCGCCCGCATCTCTATTCCGAGTATGA 411

RESULT 6

B200714

LOCUS

DEFINITION

B200714

ACCESSION

B200714.2

KEYWORDS

EST.

SOURCE

Mus musculus

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Kanaoka,T.,

Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J.,

Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K.,

Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K.,

Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,

Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,

Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.

RIKEN Mouse ESTs (Arakawa,T., et al. 2001)

Unpublished (2001)

On Jun 30, 2000 this sequence version replaced gi:8865667.

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@gsr.riken.jp, URL: <http://genome.gsc.riken.jp/>

COMMENT

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. *Genome Res.* 10 (10), 1617-1630 (2000)

wagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. *Genome Res.* 10 (11), 1757-1771 (2000)

Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y. and Hayashizaki, Y.

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. *Genome Res.* 11 (2), 281-289 (2001)

Kondo, S., Shinagawa, A., Saito, T., Kiyosawa, H., Yamanaka, I., Aizawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and Hayashizaki, Y.

Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences *Mamm. Genome.* 12, 673-677 (2001)

Please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

FEATURES

Location/Qualifiers

```
1..627
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="A430027L01"
/tissue_type="thymus"
/dev_stage="0 day neonate"
/lab_host="DH10B"
/clone_lib="RIKEN full-length enriched, 0 day neonate thymus"
/note="Site 1: Sali; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 459.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCGAGTTAATTAATTCGAGATG 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC 1."
```

ORIGIN

```
Query Match 76.6%; Score 31.4; DB 2; Length 627;
Best Local Similarity 85.4%; Pred. No. 0.063;
Matches 35; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 TTGTTGATGCTGAGCCCCCACTCTATTCGAGTATGA 41
|||||
170 TTGTTGATGCTGAGCCGCCATGATCTATTCTGAATATGA 210
```

RESULT 7

```
AK039911
LOCUS 2109 bp mRNA linear HTC 03-APR-2004
DEFINITION Mus musculus 0 day neonate thymus cDNA, RIKEN full-length enriched library, clone:A430027L01 product:estrogen receptor 1 (alpha), full
```

insert sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AK039911 GI:26087489

HTC; CAP trapper.

Mus musculus (house mouse)

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS

Carninci, P. and Hayashizaki, Y.

High-efficiency full-length cDNA cloning

Meth. Enzymol. 303, 19-44 (1999)

JOURNAL
MEDLINE
PUBMED

99279253
10349636

REFERENCE
AUTHORS

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes *Genome Res.* 10 (10), 1617-1630 (2000)

JOURNAL
MEDLINE
PUBMED

20499374
11042159

REFERENCE
AUTHORS

Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsumi, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer

Genome Res. 10 (11), 1757-1771 (2000)

JOURNAL
MEDLINE
PUBMED

20530913
11076861

REFERENCE
AUTHORS

The RIKEN Genome Exploration Research Group Phase II Team and the PANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)

JOURNAL
REFERENCE
AUTHORS

The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)

JOURNAL
REFERENCE
AUTHORS

Adachi, U., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, K., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.

JOURNAL
REFERENCE
AUTHORS

Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@sc.riken.jp, URL: <http://genome.gsc.riken.jp/>, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

JOURNAL
REFERENCE
AUTHORS

Direct Submission
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@sc.riken.jp, URL: <http://genome.gsc.riken.jp/>, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

JOURNAL
REFERENCE
AUTHORS

Direct Submission
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@sc.riken.jp, URL: <http://genome.gsc.riken.jp/>, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

JOURNAL
REFERENCE
AUTHORS

Direct Submission
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@sc.riken.jp, URL: <http://genome.gsc.riken.jp/>, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

JOURNAL
REFERENCE
AUTHORS

Direct Submission
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@sc.riken.jp, URL: <http://genome.gsc.riken.jp/>, Tel: 81-45-503-9222, Fax: 81-45-503-9216)


```

URL:http://genome.gsc.riken.jp/
URL:http://fantom.gsc.riken.jp/.
FEATURES
    source
        1..2109
            /location/Qualifiers
            /organism="Mus musculus"
            /mol_type="mRNA"
            /strain="C57BL/6J"
            /db_xref="FANTOM DB:A430027L01"
            /db_xref="taxon:10090"
            /clone="A430027L01"
            /tissue_type="thymus"
            /clone_lib="RIKEN full-length enriched mouse cDNA library"
            /dev_stage="0 day neonate"
            1..2109
            /note="estrogen receptor 1 (alpha) (MGD|GCI:1352467,
            GB|NM_007956, evidence: BLASTN, 99%, match=1960)"
        misc_feature
            1..2109
            /note="estrogen receptor 1 (alpha) (MGD|GCI:1352467,
            GB|NM_007956, evidence: BLASTN, 99%, match=1960)"
        ORIGIN
            Query Match 76.6%; Score 31.4; DB 3; Length 2109;
            Best Local Similarity 85.4%; Pred. No. 0.071;
            Matches 35; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 TTGTTGGATGCTGACCCCGCCCTACTCTATTCGAGATGA 41
    |||||
Db 1652 TTGTTGGATGCTGAACCGCCCGCCCTACTCTATTCGAGATGA 1692

RESULT 8
AK077236
LOCUS
DEFINITION
    Mus musculus 11 days pregnant adult female ovary and uterus cDNA,
    RIKEN full-length enriched library, clone:5031409B20
    product:estrogen receptor 1 (alpha), full insert sequence.
ACCESSION
    AK077236
VERSION
    AK077236 1 GI:26097248
KEYWORDS
    HTC; CAP trapper.
SOURCE
    Mus musculus (house mouse)
    ORGANISM
        Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
        Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
        1
REFERENCE
    Carninci,P. and Hayashizaki,Y.
    High-efficiency full-length cDNA cloning
    Meth. Enzymol. 303, 19-44 (1999)
    99279253
    PUBMED
    10349636
    2
REFERENCE
    Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
    Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
    Normalization and subtraction of cap-trapper-selected cDNAs to
    prepare full-length cDNA libraries for rapid discovery of new genes
    Genome Res. 10 (10), 1617-1630 (2000)
    20499374
    PUBMED
    11042159
    3
REFERENCE
    Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
    Konno,H., Akiyama,J., Nishi,K., Kitsuai,T., Tashiro,H., Itoh,M.,
    Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishire,T., Harada,A.,
    Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
    Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
    Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J.,
    Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
    RIKEN integrated sequence analysis (RISA) system--384-format
    sequencing pipeline with 384 multicapillary sequencer
    Genome Res. 10 (11), 1757-1771 (2000)
    20530913
    PUBMED
    11076861
    4
REFERENCE
    The RIKEN Genome Exploration Research Group Phase II Team and the
    FANTOM Consortium.
    Functional annotation of a full-length mouse cDNA collection
    Nature 409, 685-690 (2001)
    5

```

```

AUTHORS
    The FANTOM Consortium and the RIKEN Genome Exploration Research
    Group Phase I & II Team.
    Analysis of the mouse transcriptome based on functional annotation
    of 60,770 full-length cDNAs
    Nature 420, 563-573 (2002)
    6 (bases 1 to 2591)
    Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
    Fukuda,S., Furuno,M., Hanagaki,T., Harai,A., Hashizume,W.,
    Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozawa,T.,
    Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
    Kato,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
    Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
    Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
    Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
    Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
    Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,P., Takaku-Akahira,S.,
    Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
    Muramatsu,M. and Hayashizaki,Y.
    Direct Submission
    Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
    Physical and Chemical Research (RIKEN), Laboratory for Genome
    Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
    RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
    Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.jp,
    URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222,
    Fax:81-45-503-9216)
    cDNA library was prepared and sequenced in Mouse Genome
    Encyclopedia Project of Genome Exploration Research Group in Riken
    Genomic Sciences Center and Genome Science Laboratory in RIKEN.
    Division of Experimental Animal Research in Riken contributed to
    prepare mouse tissues.
    Please visit our web site for further details.
    URL:http://genome.gsc.riken.jp/
    URL:http://fantom.gsc.riken.jp/.
    Location/Qualifiers
        1..2591
            /organism="Mus musculus"
            /mol_type="mRNA"
            /strain="C57BL/6J"
            /db_xref="FANTOM DB:5031409B20"
            /db_xref="taxon:10090"
            /clone="5031409B20"
            /sex="female"
            /tissue_type="ovary and uterus"
            /clone_lib="RIKEN full-length enriched mouse cDNA library"
            /dev_stage="11 days pregnant adult"
            188..1988
            /note="estrogen receptor 1 (alpha) (MGD|GCI:1352467,
            GB|NM_007956, evidence: BLASTN, 99%, match=1960)
            putative"
        misc_feature
            Query Match 76.6%; Score 31.4; DB 3; Length 2591;
            Best Local Similarity 85.4%; Pred. No. 0.072;
            Matches 35; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 TTGTTGATGCTGAGCCCCCCTACTCTATTCGAGATGA 41
    |||||
Db 1154 TTGTTGATGCTGAACCGCCCGCCCTACTCTATTCGAGATGA 1194

RESULT 9
AK054182
LOCUS
DEFINITION
    Mus musculus 2 days pregnant adult female oviduct cDNA, RIKEN
    full-length enriched library, clone:E230025M07 product:estrogen
    receptor 1 (alpha), full insert sequence.
ACCESSION
    AK054182
VERSION
    AK054182.1 GI:26096219
KEYWORDS
    HTC; CAP trapper.
SOURCE
    Mus musculus (house mouse)
    ORGANISM
        Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```



```

REFERENCE
AUTHORS
6 (bases 1 to 4248)
Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
Hori,F., Imorani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
Kato,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
Muramatsu,M. and Hayashizaki,Y.
Direct Submission
TITLE
JOURNAL
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.jp,
URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222,
Fax:81-45-503-9216)
COMMENT
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Tissues were provided by Dr. John Todd (Dept. of Medical Genetics
Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome
Trust/MRC building Addenbrookes Hospital Cambridge) whose
assistance we gratefully acknowledge.
Please visit our web site for further details.
URL:http://genome.gsc.riken.jp/
URL:http://fantom.gsc.riken.jp/.
Location/Qualifiers
1..4248
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:A630019F10"
/db_xref="taxon:10090"
/clone="A630019F10"
/tissue_type="thymus"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="3 days neonate"
331..2130
/note="unnamed protein product; estrogen receptor 1
(alpha) (MGD|MG1:1352467, GB|NM_007956, evidence: BLASTN,
99%, match=1960)
putative"
CDS
/codon_start=1
/protein_id="BAC30973.1"
/db_xref="GI:26334545"
/translation="MTMTLTKSGMALLHQIQNELEPLNRQLKMPERALGEVTV
DNSKPTVFNYPGEAAVEFAAARAAASAPVYQSGIAYPGSEAAAFSANSLGAPP
QLNSVSPFLMLHPQLSPFLPHGQVQVYLENEPSAYAVRDTPGFAYFSNDSN
RQNGRERLSSNKGNMESAKETRYCAVNDYASGYGVSGCEGCKAFKRSIQ
GHNDYCPATNQCTIDKNRKSQACRLKCYEGVMGKGIKRDGRGMLKHKRQD
DLGRNEMASGDMRAANLWPSPLVIKHTKNSPALSDADQVMSALLDAEPWYSE
YDPSRPESEASMGGLTNLADRELVMINWAKRVPFGDNLNHDQVHLLDECAWLEILM
IGLWRSMEHPGKLLFAPNLLDRNCKQCVGEWIFDMLLATSRRFMNMQGEEFV
AQLLILSHRHSNKGMEHLNMYNCKNVVPLVDLLEMLDHLHAPASRMGVPEE
PSQTQLATTSSTSAHSIQTYVIPPEAGFENFI"
ORIGIN
Query Match 76.6%; Score 31.4; DB 3; Length 4248;
Best Local Similarity 85.4%; Pred. No. 0.075;
Matches 35; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
1 TTGTTGATGCTGAGCCGCCCATCTCTATTCGAGTATGA 41
2997 TTTGTTGATGCTGAGCCGCCCATCTCTATTCGAGTATGA 1337
2998
2999
3000
3001
3002
3003
3004
3005
3006
3007
3008
3009
3010
3011
3012
3013
3014
3015
3016
3017
3018
3019
3020
3021
3022
3023
3024
3025
3026
3027
3028
3029
3030
3031
3032
3033
3034
3035
3036
3037
3038
3039
3040
3041
3042
3043
3044
3045
3046
3047
3048
3049
3050
3051
3052
3053
3054
3055
3056
3057
3058
3059
3060
3061
3062
3063
3064
3065
3066
3067
3068
3069
3070
3071
3072
3073
3074
3075
3076
3077
3078
3079
3080
3081
3082
3083
3084
3085
3086
3087
3088
3089
3090
3091
3092
3093
3094
3095
3096
3097
3098
3099
3100
3101
3102
3103
3104
3105
3106
3107
3108
3109
3110
3111
3112
3113
3114
3115
3116
3117
3118
3119
3120
3121
3122
3123
3124
3125
3126
3127
3128
3129
3130
3131
3132
3133
3134
3135
3136
3137
3138
3139
3140
3141
3142
3143
3144
3145
3146
3147
3148
3149
3150
3151
3152
3153
3154
3155
3156
3157
3158
3159
3160
3161
3162
3163
3164
3165
3166
3167
3168
3169
3170
3171
3172
3173
3174
3175
3176
3177
3178
3179
3180
3181
3182
3183
3184
3185
3186
3187
3188
3189
3190
3191
3192
3193
3194
3195
3196
3197
3198
3199
3200
3201
3202
3203
3204
3205
3206
3207
3208
3209
3210
3211
3212
3213
3214
3215
3216
3217
3218
3219
3220
3221
3222
3223
3224
3225
3226
3227
3228
3229
3230
3231
3232
3233
3234
3235
3236
3237
3238
3239
3240
3241
3242
3243
3244
3245
3246
3247
3248
3249
3250
3251
3252
3253
3254
3255
3256
3257
3258
3259
3260
3261
3262
3263
3264
3265
3266
3267
3268
3269
3270
3271
3272
3273
3274
3275
3276
3277
3278
3279
3280
3281
3282
3283
3284
3285
3286
3287
3288
3289
3290
3291
3292
3293
3294
3295
3296
3297
3298
3299
3300
3301
3302
3303
3304
3305
3306
3307
3308
3309
3310
3311
3312
3313
3314
3315
3316
3317
3318
3319
3320
3321
3322
3323
3324
3325
3326
3327
3328
3329
3330
3331
3332
3333
3334
3335
3336
3337
3338
3339
3340
3341
3342
3343
3344
3345
3346
3347
3348
3349
3350
3351
3352
3353
3354
3355
3356
3357
3358
3359
3360
3361
3362
3363
3364
3365
3366
3367
3368
3369
3370
3371
3372
3373
3374
3375
3376
3377
3378
3379
3380
3381
3382
3383
3384
3385
3386
3387
3388
3389
3390
3391
3392
3393
3394
3395
3396
3397
3398
3399
3400
3401
3402
3403
3404
3405
3406
3407
3408
3409
3410
3411
3412
3413
3414
3415
3416
3417
3418
3419
3420
3421
3422
3423
3424
3425
3426
3427
3428
3429
3430
3431
3432
3433
3434
3435
3436
3437
3438
3439
3440
3441
3442
3443
3444
3445
3446
3447
3448
3449
3450
3451
3452
3453
3454
3455
3456
3457
3458
3459
3460
3461
3462
3463
3464
3465
3466
3467
3468
3469
3470
3471
3472
3473
3474
3475
3476
3477
3478
3479
3480
3481
3482
3483
3484
3485
3486
3487
3488
3489
3490
3491
3492
3493
3494
3495
3496
3497
3498
3499
3500
3501
3502
3503
3504
3505
3506
3507
3508
3509
3510
3511
3512
3513
3514
3515
3516
3517
3518
3519
3520
3521
3522
3523
3524
3525
3526
3527
3528
3529
3530
3531
3532
3533
3534
3535
3536
3537
3538
3539
3540
3541
3542
3543
3544
3545
3546
3547
3548
3549
3550
3551
3552
3553
3554
3555
3556
3557
3558
3559
3560
3561
3562
3563
3564
3565
3566
3567
3568
3569
3570
3571
3572
3573
3574
3575
3576
3577
3578
3579
3580
3581
3582
3583
3584
3585
3586
3587
3588
3589
3590
3591
3592
3593
3594
3595
3596
3597
3598
3599
3600
3601
3602
3603
3604
3605
3606
3607
3608
3609
3610
3611
3612
3613
3614
3615
3616
3617
3618
3619
3620
3621
3622
3623
3624
3625
3626
3627
3628
3629
3630
3631
3632
3633
3634
3635
3636
3637
3638
3639
3640
3641
3642
3643
3644
3645
3646
3647
3648
3649
3650
3651
3652
3653
3654
3655
3656
3657
3658
3659
3660
3661
3662
3663
3664
3665
3666
3667
3668
3669
3670
3671
3672
3673
3674
3675
3676
3677
3678
3679
3680
3681
3682
3683
3684
3685
3686
3687
3688
3689
3690
3691
3692
3693
3694
3695
3696
3697
3698
3699
3700
3701
3702
3703
3704
3705
3706
3707
3708
3709
3710
3711
3712
3713
3714
3715
3716
3717
3718
3719
3720
3721
3722
3723
3724
3725
3726
3727
3728
3729
3730
3731
3732
3733
3734
3735
3736
3737
3738
3739
3740
3741
3742
3743
3744
3745
3746
3747
3748
3749
3750
3751
3752
3753
3754
3755
3756
3757
3758
3759
3760
3761
3762
3763
3764
3765
3766
3767
3768
3769
3770
3771
3772
3773
3774
3775
3776
3777
3778
3779
3780
3781
3782
3783
3784
3785
3786
3787
3788
3789
3790
3791
3792
3793
3794
3795
3796
3797
3798
3799
3800
3801
3802
3803
3804
3805
3806
3807
3808
3809
3810
3811
3812
3813
3814
3815
3816
3817
3818
3819
3820
3821
3822
3823
3824
3825
3826
3827
3828
3829
3830
3831
3832
3833
3834
3835
3836
3837
3838
3839
3840
3841
3842
3843
3844
3845
3846
3847
3848
3849
3850
3851
3852
3853
3854
3855
3856
3857
3858
3859
3860
3861
3862
3863
3864
3865
3866
3867
3868
3869
3870
3871
3872
3873
3874
3875
3876
3877
3878
3879
3880
3881
3882
3883
3884
3885
3886
3887
3888
3889
3890
3891
3892
3893
3894
3895
3896
3897
3898
3899
3900
3901
3902
3903
3904
3905
3906
3907
3908
3909
3910
3911
3912
3913
3914
3915
3916
3917
3918
3919
3920
3921
3922
3923
3924
3925
3926
3927
3928
3929
3930
3931
3932
3933
3934
3935
3936
3937
3938
3939
3940
3941
3942
3943
3944
3945
3946
3947
3948
3949
3950
3951
3952
3953
3954
3955
3956
3957
3958
3959
3960
3961
3962
3963
3964
3965
3966
3967
3968
3969
3970
3971
3972
3973
3974
3975
3976
3977
3978
3979
3980
3981
3982
3983
3984
3985
3986
3987
3988
3989
3990
3991
3992
3993
3994
3995
3996
3997
3998
3999
4000
4001
4002
4003
4004
4005
4006
4007
4008
4009
4010
4011
4012
4013
4014
4015
4016
4017
4018
4019
4020
4021
4022
4023
4024
4025
4026
4027
4028
4029
4030
4031
4032
4033
4034
4035
4036
4037
4038
4039
4040
4041
4042
4043
4044
4045
4046
4047
4048
4049
4050
4051
4052
4053
4054
4055
4056
4057
4058
4059
4060
4061
4062
4063
4064
4065
4066
4067
4068
4069
4070
4071
4072
4073
4074
4075
4076
4077
4078
4079
4080
4081
4082
4083
4084
4085
4086
4087
4088
4089
4090
4091
4092
4093
4094
4095
4096
4097
4098
4099
4100
4101
4102
4103
4104
4105
4106
4107
4108
4109
4110
4111
4112
4113
4114
4115
4116
4117
4118
4119
4120
4121
4122
4123
4124
4125
4126
4127
4128
4129
4130
4131
4132
4133
4134
4135
4136
4137
4138
4139
4140
4141
4142
4143
4144
4145
4146
4147
4148
4149
4150
4151
4152
4153
4154
4155
4156
4157
4158
4159
4160
4161
4162
4163
4164
4165
4166
4167
4168
4169
4170
4171
4172
4173
4174
4175
4176
4177
4178
4179
4180
4181
4182
4183
4184
4185
4186
4187
4188
4189
4190
4191
4192
4193
4194
4195
4196
4197
4198
4199
4200
4201
4202
4203
4204
4205
4206
4207
4208
4209
4210
4211
4212
4213
4214
4215
4216
4217
4218
4219
4220
4221
4222
4223
4224
4225
4226
4227
4228
4229
4230
4231
4232
4233
4234
4235
4236
4237
4238
4239
4240
4241
4242
4243
4244
4245
4246
4247
4248
4249
4250
4251
4252
4253
4254
4255
4256
4257
4258
4259
4260
4261
4262
4263
4264
4265
4266
4267
4268
4269
4270
4271
4272
4273
4274
4275
4276
4277
4278
4279
4280
4281
4282
4283
4284
4285
4286
4287
4288
4289
4290
4291
4292
4293
4294
4295
4296
4297
4298
4299
4300
4301
4302
4303
4304
4305
4306
4307
4308
4309
4310
4311
4312
4313
4314
4315
4316
4317
4318
4319
4320
4321
4322
4323
4324
4325
4326
4327
4328
4329
4330
4331
4332
4333
4334
4335
4336
4337
4338
4339
4340
4341
4342
4343
4344
4345
4346
4347
4348
4349
4350
4351
4352
4353
4354
4355
4356
4357
4358
4359
4360
4361
4362
4363
4364
4365
4366
4367
4368
4369
4370
4371
4372
4373
4374
4375
4376
4377
4378
4379
4380
4381
4382
4383
4384
4385
4386
4387
4388
4389
4390
4391
4392
4393
4394
4395
4396
4397
4398
4399
4400
4401
4402
4403
4404
4405
4406
4407
4408
4409
4410
4411
4412
4413
4414
4415
4416
4417
4418
4419
4420
4421
4422
4423
4424
4425
4426
4427
4428
4429
4430
4431
4432
4433
4434
4435
4436
4437
4438
4439
4440
4441
4442
4443
4444
4445
4446
4447
4448
4449
4450
4451
4452
4453
4454
4455
4456
4457
4458
4459
4460
4461
4462
4463
4464
4465
4466
4467
4468
4469
4470
4471
4472
4473
4474
4475
4476
4477
4478
4479
4480
4481
4482
4483
4484
4485
4486
4487
4488
4489
4490
4491
4492
4493
4494
4495
4496
4497
4498
4499
4500
4501
4502
4503
4504
4505
4506
4507
4508
4509
4510
4511
4512
4513
4514
4515
4516
4517
4518
4519
4520
4521
4522
4523
4524
4525
4526
4527
4528
4529
4530
4531
4532
4533
4534
4535
4536
4537
4538
4539
4540
4541
4542
4543
4544
4545
4546
4547
4548
4549
4550
4551
4552
4553
4554
4555
4556
4557
4558
4559
4560
4561
4562
4563
4564
4565
4566
4567
4568
4569
4570
4571
4572
4573
4574
4575
4576
4577
4578
4579
4580
4581
4582
4583
4584
4585
4586
4587
4588
4589
4590
4591
4592
4593
4594
4595
4596
4597
4598
4599
4600
4601
4602
4603
4604
4605
4606
4607
4608
4609
4610
4611
4612
4613
4614
4615
4616
4617
4618
4619
4620
4621
4622
4623
4624
4625
4626
4627
4628
4629
4630
4631
4632
4633
4634
4635
4636
4637
4638
4639
4640
4641
4642
4643
4644
4645
4646
4647
4648
4649
4650
4651
4652
4653
4654
4655
4656
4657
4658
4659
4660
4661
4662
4663
4664
4665
4666
4667
4668
4669
4670
4671
4672
4673
4674
4675
4676
4677
4678
4679
4680
4681
4682
4683
4684
4685
4686
4687
4688
4689
4690
4691
4692
4693
4694
4695
4696
4697
4698
4699
4700
4701
4702
4703
4704
4705
4706
4707
4708
4709
4710
4711
4712
4713
4714
4715
4716
4717
4718
4719
4720
4721
4722
4723
4724
4725
4726
4727
4728
4729
4730
4731
4732
4733
4734
4735
4736
4737
4738
4739
4740
4741
4742
4743
4744
4745
4746
4747
4748
4749
4750
4751
4752
4753
4754
4755
4756
4757
4758
4759
4760
4761
4762
4763
4764
4765
4766
4767
4768
4769
4770
4771
4772
4773
4774
4775
4776
4777
4778
4779
4780
4781
4782
4783
4784
4785
4786
4787
4788
4789
4790
4791
4792
4793
4794
4795
4796
4797
4798
4799
4800
4801
4802
4803
4804
4805
4806
4807
4808
4809
4810
4811
4812
4813
4814
4815
4816
4817
4818
4819
4820
4821
4822
4823
4824
4825
4826
4827
4828
4829
4830
4831
4832
4833
4834
4835
4836
4837
4838
4839
4840
4841
4842
4843
4844
4845
4846
4847
4848
4849
4850
4851
4852
4853
4854
4855
4856
4857
4858
4859
4860
4861
4862
4863
4864
4865
4866
4867
4868
4869
4870
4871
4872
4873
4874
4875
4876
4877
4878
4879
4880
4881
4882
4883
4884
4885
4886
4887
4888
4889
4890
4891
4892
4893
4894
4895
4896
4897
4898
4899
4900
4901
4902
4903
4904
4905
4906
4907
4908
4909
4910
4911
4912
4913
4914
4915
4916
4917
4918
4919
4920
4921
4922
4923
4924
4925
4926
4927
4928
4929
4930
4931
4932
4933
4934
4935
4936
4937
4938
4939
4940
4941
4942
4943
4944
4945
4946
4947
4948
4949
4950
4951
4952
4953
4954
4955
4956
4957
4958
4959
4960
4961
4962
4963
4964
4965
4966
4967
4968
4969
4970
4971
4972
4973
4974
4975
4976
4977
4978
4979
4980
4981
4982
4983
4984
4985
4986
4987
4988
4989
4990
4991
4992
4993
4994
4995
4996
4997
4998
4999
5000
5001
5002
5003
5004
5005
5006
5007
5008
5009
5010
5011
5012
5013
5014
5015
5016
5017
5018
5019
5020
5021
5022
5023
5024
5025
5026
5027
5028
5029
5030
5031
5032
5033
5034
5035
5036
5037
5038
5039
5040
5041
5042
5043
5044
5045
5046
5047
5048
5049
5050
5051
5052
5053
5054
5055
5056
5057
5058
5059
5060
5061
5062
5063
5064
5065
5066
5067
5068
5069
5070
5071
5072
5073
5074
5075
5076
5077
5078
5079
5080
5081
5082
5083
5084
5085
5086
5087
5088
5089
5090
5091
5092
5093
5094
5095
5096
5097
5098
5099
5100
5101
5102
5103
5104
5105
5106
5107
5108
5109
5110
5111
5112
5113
5114
5115
5116
5117
5118
5119
5120
5121
5122
5123
5124
5125
5126
5127
5128
5129
5130
5131
5132
5133
5134
5135
5136
5137
5138
5139
5140
5141
5142
5143
5144
5145
5146
5147
5148
5149
5150
5151
5152
5153
5154
5155
5156
5157
5158
5159
5160
5161
5162
5163
5164
5165
5166
5167
5168
5169
5170
5171
5172
5173
5174
5175
5176
5177
5178
5179
5180
5181
5182
518
```

COMMENT	cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. URL:http://genome.gsc.riken.jp/ URL:http://fantom.gsc.riken.jp/.		
FEATURES	source	Location/Qualifiers 1..4315 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6J" /db_xref="FANTOM DB:E230030D02" /db_xref="taxon:10090" /clone="E230030D02" /sex="female" /tissue_type="oviduct" /clone_lib="RIKEN full-length enriched mouse cDNA library" /dev_stage="2 days pregnant adult" misc_feature 1..4315 /note="estrogen receptor 1 (alpha) (MGJ MGJ:1352467, GB NM_007956, evidence: BLASTN, 99%, match=1960)"	
ORIGIN	Query Match Best Local Similarity Matches 35; Conservative	76.6%; Score 31.4; DB 3; Length 4315; 85.4%; Pred. No. 0.076; 6; Indels 0; Gaps 0; 0; Mismatches 6;	
QY	1	TTGTTGGATGCTGAGCCGCCATCTCTATTCGAGTATGA	41
Db	1129	TTGTTGGATGCTGAACCCGCCATGATCTATTCTGAATATGA	1169
RESULT 12	AK036627		
LOCUS	Mus musculus adult male bone cDNA, RIKEN full-length enriched library, clone:9830143G19 product:estrogen receptor 1 (alpha), full insert sequence.		
DEFINITION	AK036627		
ACCESSION	AK036627.1 GI:26331559		
VERSION	HTC; CAP trapper.		
KEYWORDS	Mus musculus (house mouse)		
SOURCE	Mus musculus		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
REFERENCE	1		
AUTHORS	Carninci, P. and Hayashizaki, Y.		
TITLE	High-efficiency full-length cDNA cloning		
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)		
MEDLINE	99279253		
PUBMED	10349636		
REFERENCE	2		
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.		
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes		
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)		
MEDLINE	20499374		
PUBMED	11042159		
REFERENCE	3		
AUTHORS	Shibata, K., Itoh, M., Aizawa, K., Nagao, K., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsu, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.		
TITLE	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer		
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)		
MEDLINE	20530913		
REFERENCE	4		
AUTHORS	The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.		
TITLE	Functional annotation of a full-length mouse cDNA collection		
JOURNAL	Nature 409, 685-690 (2001)		
MEDLINE	11076861		
PUBMED	11076861		
REFERENCE	5		
AUTHORS	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.		
TITLE	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs		
JOURNAL	Nature 420, 563-573 (2002)		
MEDLINE	6 (bases 1 to 4321)		
PUBMED	Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.		
ORIGIN	Direct Submission	Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@sc.riken.jp, URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)	
COMMENT	CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Please visit our web site for further details. URL:http://genome.gsc.riken.jp/ URL:http://fantom.gsc.riken.jp/.		
FEATURES	source	Location/Qualifiers 1..4321 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6J" /db_xref="FANTOM DB:9830143G19" /db_xref="taxon:10090" /clone="9830143G19" /sex="male" /tissue_type="bone" /clone_lib="RIKEN full-length enriched mouse cDNA library" /dev_stage="adult" 183..1982 /note="unnamed protein product; estrogen receptor 1 (alpha) (MGJ MGJ:1352467, GB NM_007956, evidence: BLASTN, 99%, match=1960) putative" /codon_start=1 /protein_id="BAC29510.1" /db_xref="GI:26331560" /translation="MTMTLTKASGMALLHQIQNELEPLNRLPKMPMERALGEVYV DSKPTVFNYPPEGAAYEFNAHAAAAASAPVGGSIAYGFGSEAAFSANSLGAPP QLNVSFSPFLMLHPPLSPFLPHGQVVPVLENEPSAYAVRDTGPAFYRSNDN RQNCRESLSNNEKGMTESAKTRYCAVNDYASGVYGVWCEGCKAFKRSIQ GNDYCPATNCTIDNKRKSCQACRLKCYGVNMGKGRKDRGRLMKHQRQD DLEGRNEMASGMDRAANLWPSPLVTKHKYKNSPALSLTADQVSLDLAEPMYSE YDFRFSFASMMGLTNLADRELHMINWAKRVPFGDNLNLDQVHLLCAWLEILM IGLVWRSMHPGKLLFAPNLLDRNCKVEGVMEIFDMLLTSRFRMNLQOHRRL CLKSIILLNSGVYVTFLLSKLSLEEDHIVLDKTDITLHLMKAGLTQOHRRL AQOLLTLSHIRHNSKGMHLYNMCKNVPLVLLLEMLDHLRLHAPASRNGVPPEE PSQTQATTSSTSAHSLQTYIYPEAEGFPNTI"	
	polyA_signal	4306..4311	

```

polyA_site
/note="putative"
4321
/note="putative"

ORIGIN
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 576)
Ko.M.S.H., Kitchen,J.R., Wang,X., Threat,T.A., Sun,T.,
Depalma,G.E., Liang,Y., Kargul,G.J., Sharara,R., Lim,M.K. and
Doi,H.
Systematic analyses of genes expressed in 4-cell mouse embryo (The
ERATO/Doi Project at Wayne State University)
Unpublished (1998)
Contact: Hirofumi Doi
Doi Biosymmetry Project, ERATO
Japan Science and Technology Corporation (JST)
WBG Marine East 12F, 2-6 Nakase, Mihama-ku, Chiba 261-71, Japan
Email: hdo@bioa.jst.go.jp.
Location/Qualifiers
1..576
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="J0828F12"
/dev_stage="four-cell-embryo"
/clone_lib="Mouse four-cell-embryo cDNA"

ORIGIN
Query Match 70.2%; Score 28.8; DB 1; Length 576;
Best Local Similarity 80.5%; Pred. No. 0.72;
Matches 33; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 TTGTGGATGCTGAGCCGCCCATCTCTATTCCGAGTATGA 41
Db 460 TTGTGGATGCTGAGCCGCCCATCTCTATTCCGAGTATGA 420

RESULT 15
CG483529
LOCUS
DEFINITION
CG483529 656 bp DNA linear GSS 01-OCT-2003
OST17199 Mus musculus 129Sv/Ev Mus musculus genomic clone OST17199,
genomic survey sequence.
CG483529
ACCESSION
CG483529.1 GI:37238317
VERSION
KEYWORDS
SOURCE
Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 656)
Zambrowicz,B.P., Abuin,A., Ramirez-Solis,R., Richter,L.J.,
Piggott,J., BeltrandelRio,H., Buxton,E.C., Edwards,J., Finch,R.A.,
Fridde,C.J., Gupta,A., Hansen,G., Hu,Y., Huang,W., Jaing,C.,
Key,B.W. Jr., Kipp,P., Kohlhauff,B., Ma,Z.-Q., Markesich,D.,
Payne,R., Potter,D.G., Qian,N., Shaw,J., Schrick,J., Shi,Z.-Z.,
Sparks,M.J., Van Slichtenhorst,I., Vogel,P., Walke,W., Xu,N.,
Zhu,Q., Person,C. and Sands,A.T.
Wnk1 kinase deficiency lowers blood pressure in mice: a gene-trap
screen to identify potential targets for therapeutic intervention
Proc. Natl. Acad. Sci. U.S.A. 100 (24), 14109-14114 (2003)
Contact: Zambrowicz BP
OmniBank
Lexicon Genetics Incorporated
4000 Research Forest Drive, The Woodlands, TX 77381, USA
Email: materials@lexgen.com
Gene trap sequence tag generated by 3' RACE from mouse ES cells as
described in Zambrowicz et al (Nature. 1998 Apr 9;392(6676):608-11)
Class: Gene Trap.
Location/Qualifiers
1..656
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="129Sv/Ev"
/db_xref="taxon:10090"

FEATURES
source
QY 1 TTGTGGATGCTGAGCCGCCCATCTCTATTCCGAGTATGA 41
Db 1144 TTGTGGATGCTGAGCCGCCCATCTCTATTCCGAGTATGA 1189

RESULT 13
AA092184
LOCUS
DEFINITION
114609, seq.F Human fetal heart, Lambda ZAP Express Homo sapiens
cDNA 5', mRNA sequence.
ACCESSION
AA092184
VERSION
AA092184.1 GI:1637173
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 365)
Liew,C.C.
cDNAs from fetal heart (1996)
Unpublished (1996)
Contact: Liew CC
Brigham and Women's Hospital
Harvard Medical School
75 Francis St. Boston, MA 02115, USA
Tel: 6177328915
Fax: 6179750995
Email: cliw@rics.bwh.harvard.edu
PCR Primers
FORWARD: 5' GCCAAGTCGAATTAACCCCTCACTAAAGG 3'
BACKWARD: 5' CCAGTGAATTGTAACGACTCACTATAGGCG 3'
Seq primer: 5' GAAATTAACCCCTCACTAAAGG 3'.
Location/Qualifiers
1..365
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone_lib="E. coli XL1-Blue"
/notes="Vector: Lambda ZAP Express; Site 1: EcoRI; Site 2:
XhoI; mRNA was purified from human fetal hearts (8-10
weeks). cDNA was synthesized using a XhoI-Oligo dT
adaptor-primer. EcoRI adaptors were ligated, followed by
digestion with XhoI, for directional cloning into
predigested lambda ZAP Express."

ORIGIN
Query Match 70.7%; Score 29; DB 1; Length 365;
Best Local Similarity 97.6%; Pred. No. 0.57;
Matches 40; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 TTGTGGATGCTGAGCCGCCCATCTCTATTCCGAGTATGA 41
Db 114 TTGTGGATGCTGAGCCGCCCATCTCTATTCCGAGTATGA 153

RESULT 14
AA041270/c
LOCUS
DEFINITION
AU041270 Mouse four-cell-embryo cDNA Mus musculus cDNA clone
J0828F12 3', mRNA sequence.
ACCESSION
AU041270
VERSION
AU041270.1 GI:3955505
KEYWORDS
EST.

```

/clone="OST17199"
/cell_type="embryonic stem cell"
/clone_lib="Mus musculus 129Sv/Ev"

ORIGIN

Query Match 67.8%; Score 27.8; DB 9; Length 656;
Best Local Similarity 80.0%; Pred. No. 1.9;
Matches 32; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Oy 1 TTGTTGGATGCTGAGCCCCCATCTCTATTCCGAGTATG 40
|||||
Db 457 TTGTGATGCTGAGACCGCCCATGATCTTCTGATATG 496
|||||

Search completed: November 3, 2004, 09:43:33
Job time : 712.123 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 3, 2004, 06:51:04 ; Search time 74.6435 Seconds
(without alignments)
2816.475 Million cell updates/sec

Title: US-09-933-267A-1_COPY_306362_306402

Perfect score: 41

Sequence: 1 ttgttgatgctgagccccc.....atactctatccgagtatga 41

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3413475 seqs, 2563800928 residues

Total number of hits satisfying chosen parameters: 6826950

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCT NEW PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW PUB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq:*
- 17: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq:*
- 18: /cgn2_6/ptodata/2/pubpna/US10_NEW PUB.seq:*
- 19: /cgn2_6/ptodata/2/pubpna/US11_NEW PUB.seq:*
- 20: /cgn2_6/ptodata/2/pubpna/US60_NEW PUB.seq:*
- 21: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	41	100.0	392	14	US-10-052-092-4
2	41	100.0	392	15	US-10-437-107-4
3	41	100.0	1218	15	US-10-027-983-98
4	41	100.0	1218	15	US-10-448-753-98
5	41	100.0	1237	14	US-10-052-092-3
6	41	100.0	1237	15	US-10-437-107-3
7	41	100.0	1245	15	US-10-157-899A-1
8	41	100.0	1245	15	US-10-157-899A-5
9	41	100.0	1245	15	US-10-157-899A-7
10	41	100.0	1245	15	US-10-157-899A-9
11	41	100.0	1317	15	US-10-157-899A-3
12	41	100.0	1317	15	US-10-157-899A-11
					Sequence 4, Appli
					Sequence 98, Appli
					Sequence 98, Appli
					Sequence 3, Appli
					Sequence 1, Appli
					Sequence 5, Appli
					Sequence 7, Appli
					Sequence 9, Appli
					Sequence 3, Appli
					Sequence 11, Appli

13	41	100.0	1317	15	US-10-157-899A-13	Sequence 13, Appli
14	41	100.0	1317	15	US-10-157-899A-15	Sequence 15, Appli
15	41	100.0	1542	14	US-10-006-760-18	Sequence 18, Appli
16	41	100.0	1644	14	US-10-052-092-2	Sequence 2, Appli
17	41	100.0	1644	15	US-10-437-107-2	Sequence 2, Appli
18	41	100.0	1785	15	US-10-095-373A-1	Sequence 1, Appli
19	41	100.0	1788	9	US-09-853-033-1	Sequence 1, Appli
20	41	100.0	1983	9	US-09-853-033-3	Sequence 3, Appli
21	41	100.0	1983	9	US-09-853-033-5	Sequence 5, Appli
22	41	100.0	1983	9	US-09-853-033-7	Sequence 7, Appli
23	41	100.0	2028	15	US-10-095-373A-72	Sequence 72, Appli
24	41	100.0	2031	15	US-10-095-373A-68	Sequence 68, Appli
25	41	100.0	2092	10	US-09-952-680A-9	Sequence 9, Appli
26	41	100.0	2092	14	US-10-052-092-6	Sequence 6, Appli
27	41	100.0	2092	15	US-10-157-899A-54	Sequence 54, Appli
28	41	100.0	2092	15	US-10-437-107-6	Sequence 6, Appli
29	41	100.0	2106	14	US-10-052-092-5	Sequence 5, Appli
30	41	100.0	2106	15	US-10-437-107-5	Sequence 5, Appli
31	41	100.0	2178	15	US-10-095-373A-70	Sequence 70, Appli
32	41	100.0	2181	15	US-10-095-373A-66	Sequence 66, Appli
33	41	100.0	4963	9	US-09-281-674-9	Sequence 9, Appli
34	41	100.0	4963	9	US-09-777-317-9	Sequence 9, Appli
35	41	100.0	4963	9	US-09-892-227-9	Sequence 9, Appli
36	41	100.0	6450	13	US-10-096-710-2	Sequence 2, Appli
37	41	100.0	6450	13	US-10-081-563-1	Sequence 1, Appli
38	41	100.0	6450	14	US-10-052-092-1	Sequence 1, Appli
39	41	100.0	6450	14	US-10-052-092-7	Sequence 7, Appli
40	41	100.0	6450	15	US-10-207-655-60	Sequence 60, Appli
41	41	100.0	6450	15	US-10-177-293-127	Sequence 127, Appli
42	41	100.0	6450	15	US-10-027-983-3	Sequence 3, Appli
43	41	100.0	6450	15	US-10-007-926A-422	Sequence 422, Appli
44	41	100.0	6450	15	US-10-437-107-1	Sequence 1, Appli
45	41	100.0	6450	15	US-10-437-107-7	Sequence 7, Appli

ALIGNMENTS

RESULT 1
US-10-052-092-4
; Sequence 4, Application US/10052092
; Publication No. US20030027778A1
; GENERAL INFORMATION:
; APPLICANT: Fuqua, Suzanne
; APPLICANT: Allred, D.
; APPLICANT: Hopp, Torsten A.
; APPLICANT: O'Connell, Peter
; TITLE OF INVENTION: Methods and Composition in Breast Cancer Diagnosis and Therapeutic
; FILE REFERENCE: P02102US2
; CURRENT FILING DATE: 2002-01-18
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: US 60/304,018
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 392
; TYPE: DNA
; ORGANISM: Human
US-10-052-092-4

Query Match 100.0%; Score 41; DB 14; Length 392;
Best Local Similarity 100.0%; Pred. No. 2e-08;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TTGTTGGATGCTGAGCCCCCATCTCTATTCGAGTATGA 41
Db 182 TTGTTGGATGCTGAGCCCCCATCTCTATTCGAGTATGA 222
RESULT 2

```
US-10-437-107-4
; Sequence 4, Application US/10437107
; Publication No. US20030186313A1
; GENERAL INFORMATION:
; APPLICANT: Fuqua, Suzanne
; APPLICANT: Allred, D.
; APPLICANT: Hopp, Torsten A.
; APPLICANT: O'Connell, Peter
; TITLE OF INVENTION: Methods and Composition in Breast Cancer Diagnosis and Therapeutic
; FILE REFERENCE: P02102US2
; CURRENT APPLICATION NUMBER: US/10/437,107
; CURRENT FILING DATE: 2003-05-13
; PRIOR APPLICATION NUMBER: US/10/052,092
; PRIOR FILING DATE: 2002-01-18
; PRIOR APPLICATION NUMBER: US 60/262,990
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: US 60/304,018
; PRIOR FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 392
; TYPE: DNA
; ORGANISM: Human
; US-10-437-107-4

Query Match 100.0%; Score 41; DB 15; Length 392;
Best Local Similarity 100.0%; Pred. No. 2e-08;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGTTGGATGCTGAGCCCCCATACTCTATTCCGAGTATGA 41
DB 182 TTGTTGGATGCTGAGCCCCCATACTCTATTCCGAGTATGA 222

RESULT 3
US-10-027-983-98
; Sequence 98, Application US/10027983
; Publication No. US20030139360A1
; GENERAL INFORMATION:
; APPLICANT: Kenneth W. Dobie
; APPLICANT: Mark P. Roach
; TITLE OF INVENTION: ANTISENSE MODULATION OF ESTROGEN RECEPTOR ALPHA EXPRESSION
; FILE REFERENCE: RTS-0340
; CURRENT APPLICATION NUMBER: US/10/027,983
; CURRENT FILING DATE: 2001-12-18
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 98
; LENGTH: 1218
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; US-10-027-983-98

Query Match 100.0%; Score 41; DB 15; Length 1218;
Best Local Similarity 100.0%; Pred. No. 2.4e-08;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGTTGGATGCTGAGCCCCCATACTCTATTCCGAGTATGA 41
DB 385 TTGTTGGATGCTGAGCCCCCATACTCTATTCCGAGTATGA 425

RESULT 4
US-10-448-753-98
; Sequence 98, Application US/10448753
; Publication No. US20030211611A1
; GENERAL INFORMATION:
; APPLICANT: Kenneth W. Dobie
; APPLICANT: Mark P. Roach
; TITLE OF INVENTION: ANTISENSE MODULATION OF ESTROGEN RECEPTOR ALPHA EXPRESSION
; FILE REFERENCE: RTS-0340
; CURRENT APPLICATION NUMBER: US/10/448,753
```

```
; CURRENT FILING DATE: 2003-05-30
; PRIOR APPLICATION NUMBER: US/10/027,983
; PRIOR FILING DATE: 2001-12-18
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 98
; LENGTH: 1218
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; US-10-448-753-98

Query Match 100.0%; Score 41; DB 15; Length 1218;
Best Local Similarity 100.0%; Pred. No. 2.4e-08;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGTTGGATGCTGAGCCCCCATACTCTATTCCGAGTATGA 41
DB 385 TTGTTGGATGCTGAGCCCCCATACTCTATTCCGAGTATGA 425

RESULT 5
US-10-052-092-3
; Sequence 3, Application US/10052092
; Publication No. US20030027778A1
; GENERAL INFORMATION:
; APPLICANT: Fuqua, Suzanne
; APPLICANT: Allred, D.
; APPLICANT: Hopp, Torsten A.
; APPLICANT: O'Connell, Peter
; TITLE OF INVENTION: Methods and Composition in Breast Cancer Diagnosis and Therapeutic
; FILE REFERENCE: P02102US2
; CURRENT APPLICATION NUMBER: US/10/052,092
; CURRENT FILING DATE: 2002-01-18
; PRIOR APPLICATION NUMBER: US 60/262,990
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: US 60/304,018
; PRIOR FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 1237
; TYPE: DNA
; ORGANISM: Human
; US-10-052-092-3

Query Match 100.0%; Score 41; DB 14; Length 1237;
Best Local Similarity 100.0%; Pred. No. 2.4e-08;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGTTGGATGCTGAGCCCCCATACTCTATTCCGAGTATGA 41
DB 544 TTGTTGGATGCTGAGCCCCCATACTCTATTCCGAGTATGA 584

RESULT 6
US-10-437-107-3
; Sequence 3, Application US/10437107
; Publication No. US20030186313A1
; GENERAL INFORMATION:
; APPLICANT: Fuqua, Suzanne
; APPLICANT: Allred, D.
; APPLICANT: Hopp, Torsten A.
; APPLICANT: O'Connell, Peter
; TITLE OF INVENTION: Methods and Composition in Breast Cancer Diagnosis and Therapeutic
; FILE REFERENCE: P02102US2
; CURRENT APPLICATION NUMBER: US/10/437,107
; CURRENT FILING DATE: 2003-05-13
; PRIOR APPLICATION NUMBER: US/10/052,092
; PRIOR FILING DATE: 2002-01-18
; PRIOR APPLICATION NUMBER: US 60/262,990
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: US 60/304,018
; PRIOR FILING DATE: 2001-07-09
```


; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 1237
; TYPE: DNA
; ORGANISM: Human
US-10-437-107-3

Query Match 100.0%; Score 41; DB 15; Length 1237;
Best Local Similarity 100.0%; Pred. No. 2.4e-08;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGTTGGATGCTGAGCCCCCATCTACTCTATTCGAGTATGA 41
|||||
Db 544 TTGTTGGATGCTGAGCCCCCATCTACTCTATTCGAGTATGA 584
|||||

RESULT 7

US-10-157-899A-1
; Sequence 1, Application US/10157899A
; Publication No. US20030143559A1
; GENERAL INFORMATION:
; APPLICANT: Bracken, Kathryn Rene
; APPLICANT: de los Angeles, Joseph Ernest
; APPLICANT: Huang, Ying
; APPLICANT: Kadan, Michael Joseph
; APPLICANT: Ksander, Gary Michael
; APPLICANT: Zerby, Dennis
; TITLE OF INVENTION: NOVEL ESTROGEN RECEPTOR LIGAND BINDING DOMAIN VARIANTS AND NOVEL
; TITLE OF INVENTION: LIGANDS AND PHARMACEUTICAL COMPOSITIONS
; FILE REFERENCE: 4-32018A
; CURRENT APPLICATION NUMBER: US/10/157,899A
; PRIOR FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 60/294,839
; PRIOR FILING DATE: 2001-05-31
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1245
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Fusion of the human-Estrogen-receptor-ligand-binding-domain and a
; OTHER INFORMATION: zinc finger array(C7)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1245)
US-10-157-899A-1

Query Match 100.0%; Score 41; DB 15; Length 1245;
Best Local Similarity 100.0%; Pred. No. 2.4e-08;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGTTGGATGCTGAGCCCCCATCTACTCTATTCGAGTATGA 41
|||||
Db 406 TTGTTGGATGCTGAGCCCCCATCTACTCTATTCGAGTATGA 446
|||||

RESULT 8

US-10-157-899A-5
; Sequence 5, Application US/10157899A
; Publication No. US20030143559A1
; GENERAL INFORMATION:
; APPLICANT: Bracken, Kathryn Rene
; APPLICANT: de los Angeles, Joseph Ernest
; APPLICANT: Huang, Ying
; APPLICANT: Kadan, Michael Joseph
; APPLICANT: Ksander, Gary Michael
; APPLICANT: Zerby, Dennis
; TITLE OF INVENTION: NOVEL ESTROGEN RECEPTOR LIGAND BINDING DOMAIN VARIANTS AND NOVEL
; TITLE OF INVENTION: LIGANDS AND PHARMACEUTICAL COMPOSITIONS
; FILE REFERENCE: 4-32018A
; CURRENT APPLICATION NUMBER: US/10/157,899A

; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 60/294,839
; PRIOR FILING DATE: 2001-05-31
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 1245
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Fusion of a mutated human-Estrogen-receptor-ligand-binding-domain
; OTHER INFORMATION: and a zinc finger array(C7)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1245)
; FEATURE:
; NAME/KEY: mutation
; LOCATION: (613)..(615)
; FEATURE:
; NAME/KEY: mutation
; LOCATION: (721)..(723)
; FEATURE:
; NAME/KEY: mutation
; LOCATION: (733)..(735)
US-10-157-899A-5

Query Match 100.0%; Score 41; DB 15; Length 1245;
Best Local Similarity 100.0%; Pred. No. 2.4e-08;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGTTGGATGCTGAGCCCCCATCTACTCTATTCGAGTATGA 41
|||||
Db 406 TTGTTGGATGCTGAGCCCCCATCTACTCTATTCGAGTATGA 446
|||||

RESULT 9

US-10-157-899A-7
; Sequence 7, Application US/10157899A
; Publication No. US20030143559A1
; GENERAL INFORMATION:
; APPLICANT: Bracken, Kathryn Rene
; APPLICANT: de los Angeles, Joseph Ernest
; APPLICANT: Huang, Ying
; APPLICANT: Kadan, Michael Joseph
; APPLICANT: Ksander, Gary Michael
; APPLICANT: Zerby, Dennis
; TITLE OF INVENTION: NOVEL ESTROGEN RECEPTOR LIGAND BINDING DOMAIN VARIANTS AND NOVEL
; TITLE OF INVENTION: LIGANDS AND PHARMACEUTICAL COMPOSITIONS
; FILE REFERENCE: 4-32018A
; CURRENT APPLICATION NUMBER: US/10/157,899A
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 60/294,839
; PRIOR FILING DATE: 2001-05-31
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 1245
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Fusion of a mutated human-Estrogen-receptor-ligand-binding-domain
; OTHER INFORMATION: and a zinc finger array(C7)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1245)
; FEATURE:
; NAME/KEY: mutation
; LOCATION: (712)..(714)
; FEATURE:
; NAME/KEY: mutation
; LOCATION: (733)..(735)
US-10-157-899A-7

```
Query Match      100.0%; Score 41; DB 15; Length 1245;
Best Local Similarity 100.0%; Pred. No. 2.4e-08;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGTTGGATGCTGAGCCCCCATACTCTATTCCGAGTATGA 41
    |||||
Db 406 TTGTTGGATGCTGAGCCCCCATACTCTATTCCGAGTATGA 446

RESULT 10
US-10-157-899A-9
; Sequence 9, Application US/10157899A
; Publication No. US20030143559A1
; GENERAL INFORMATION:
; APPLICANT: Bracken, Kathryn Rene
; APPLICANT: de los Angeles, Joseph Ernest
; APPLICANT: Huang, Ying
; APPLICANT: Kadan, Michael Joseph
; APPLICANT: Ksander, Gary Michael
; APPLICANT: Zerby, Dennis
; TITLE OF INVENTION: NOVEL ESTROGEN RECEPTOR LIGAND BINDING DOMAIN VARIANTS AND NOVEL
; FILE REFERENCE: 4-32018A
; CURRENT APPLICATION NUMBER: US/10/157,899A
; PRIOR FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 60/294,839
; PRIOR FILING DATE: 2001-05-31
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 9
; LENGTH: 1245
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Fusion of a mutated human-Estrogen-receptor-ligand-binding-domain
; OTHER INFORMATION: and a zinc finger array(C7)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1245)
; NAME/KEY: mutation
; LOCATION: (1021)..(1023)
US-10-157-899A-9

Query Match      100.0%; Score 41; DB 15; Length 1245;
Best Local Similarity 100.0%; Pred. No. 2.4e-08;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGTTGGATGCTGAGCCCCCATACTCTATTCCGAGTATGA 41
    |||||
Db 406 TTGTTGGATGCTGAGCCCCCATACTCTATTCCGAGTATGA 446

RESULT 11
US-10-157-899A-3
; Sequence 3, Application US/10157899A
; Publication No. US20030143559A1
; GENERAL INFORMATION:
; APPLICANT: Bracken, Kathryn Rene
; APPLICANT: de los Angeles, Joseph Ernest
; APPLICANT: Huang, Ying
; APPLICANT: Kadan, Michael Joseph
; APPLICANT: Ksander, Gary Michael
; APPLICANT: Zerby, Dennis
; TITLE OF INVENTION: NOVEL ESTROGEN RECEPTOR LIGAND BINDING DOMAIN VARIANTS AND NOVEL
; FILE REFERENCE: 4-32018A
; CURRENT APPLICATION NUMBER: US/10/157,899A
; PRIOR FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 60/294,839
; PRIOR FILING DATE: 2001-05-31
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 11
; LENGTH: 1317
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Fusion of a mutated human-Estrogen-receptor-ligand-binding-domain
; OTHER INFORMATION: and a zinc finger array(C7)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1317)
; NAME/KEY: mutation
; LOCATION: (685)..(687)
; NAME/KEY: mutation
; LOCATION: (793)..(795)
; NAME/KEY: mutation
; LOCATION: (805)..(807)
US-10-157-899A-11

Query Match      100.0%; Score 41; DB 15; Length 1317;
Best Local Similarity 100.0%; Pred. No. 2.4e-08;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGTTGGATGCTGAGCCCCCATACTCTATTCCGAGTATGA 41
    |||||
Db 478 TTGTTGGATGCTGAGCCCCCATACTCTATTCCGAGTATGA 518

RESULT 12
US-10-157-899A-11
; Sequence 11, Application US/10157899A
; Publication No. US20030143559A1
; GENERAL INFORMATION:
; APPLICANT: Bracken, Kathryn Rene
; APPLICANT: de los Angeles, Joseph Ernest
; APPLICANT: Huang, Ying
; APPLICANT: Kadan, Michael Joseph
; APPLICANT: Ksander, Gary Michael
; APPLICANT: Zerby, Dennis
; TITLE OF INVENTION: NOVEL ESTROGEN RECEPTOR LIGAND BINDING DOMAIN VARIANTS AND NOVEL
; FILE REFERENCE: 4-32018A
; CURRENT APPLICATION NUMBER: US/10/157,899A
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 60/294,839
; PRIOR FILING DATE: 2001-05-31
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 11
; LENGTH: 1317
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Fusion of a mutated human-Estrogen-receptor-ligand-binding-domain
; OTHER INFORMATION: and a zinc finger array(C7)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1317)
; NAME/KEY: mutation
; LOCATION: (685)..(687)
; NAME/KEY: mutation
; LOCATION: (793)..(795)
; NAME/KEY: mutation
; LOCATION: (805)..(807)
US-10-157-899A-11

Query Match      100.0%; Score 41; DB 15; Length 1317;
Best Local Similarity 100.0%; Pred. No. 2.4e-08;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGTTGGATGCTGAGCCCCCATACTCTATTCCGAGTATGA 41
    |||||
Db 478 TTGTTGGATGCTGAGCCCCCATACTCTATTCCGAGTATGA 518

RESULT 13
US-10-157-899A-13
```

```
; SEQ ID NO 3
; LENGTH: 1317
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Fusion of the human-Estrogen-receptor-ligand-binding-domain and a
; OTHER INFORMATION: zinc finger array(C7)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1317)
; LOCATION: (1)..(1317)
US-10-157-899A-3
```

```
Query Match      100.0%; Score 41; DB 15; Length 1317;
Best Local Similarity 100.0%; Pred. No. 2.4e-08;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGTTGGATGCTGAGCCCCCATACTCTATTCCGAGTATGA 41
    |||||
Db 478 TTGTTGGATGCTGAGCCCCCATACTCTATTCCGAGTATGA 518
```

```
RESULT 12
US-10-157-899A-11
; Sequence 11, Application US/10157899A
; Publication No. US20030143559A1
; GENERAL INFORMATION:
; APPLICANT: Bracken, Kathryn Rene
; APPLICANT: de los Angeles, Joseph Ernest
; APPLICANT: Huang, Ying
; APPLICANT: Kadan, Michael Joseph
; APPLICANT: Ksander, Gary Michael
; APPLICANT: Zerby, Dennis
; TITLE OF INVENTION: NOVEL ESTROGEN RECEPTOR LIGAND BINDING DOMAIN VARIANTS AND NOVEL
; FILE REFERENCE: 4-32018A
; CURRENT APPLICATION NUMBER: US/10/157,899A
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 60/294,839
; PRIOR FILING DATE: 2001-05-31
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 11
; LENGTH: 1317
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Fusion of a mutated human-Estrogen-receptor-ligand-binding-domain
; OTHER INFORMATION: and a zinc finger array(C7)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1317)
; NAME/KEY: mutation
; LOCATION: (685)..(687)
; NAME/KEY: mutation
; LOCATION: (793)..(795)
; NAME/KEY: mutation
; LOCATION: (805)..(807)
US-10-157-899A-11

Query Match      100.0%; Score 41; DB 15; Length 1317;
Best Local Similarity 100.0%; Pred. No. 2.4e-08;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGTTGGATGCTGAGCCCCCATACTCTATTCCGAGTATGA 41
    |||||
Db 478 TTGTTGGATGCTGAGCCCCCATACTCTATTCCGAGTATGA 518
```

```
RESULT 13
US-10-157-899A-13
```

```
; Sequence 13, Application US/10157899A
; Publication No. US20030143559A1
; GENERAL INFORMATION:
; APPLICANT: Bracken, Kathryn Rene
; APPLICANT: de los Angeles, Joseph Ernest
; APPLICANT: Huang, Ying
; APPLICANT: Kadan, Michael Joseph
; APPLICANT: Ksander, Gary Michael
; APPLICANT: Zerby, Dennis
; TITLE OF INVENTION: NOVEL ESTROGEN RECEPTOR LIGAND BINDING DOMAIN VARIANTS AND NOVEL
; TITLE OF INVENTION: LIGANDS AND PHARMACEUTICAL COMPOSITIONS
; FILE REFERENCE: 4-32018A
; CURRENT APPLICATION NUMBER: US/10/157,899A
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 60/294,839
; PRIOR FILING DATE: 2001-05-31
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 13
; LENGTH: 1317
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Fusion of a mutated human-Estrogen-receptor-ligand-binding-domain
; OTHER INFORMATION: and a zinc finger array (C7)
; NAME/KEY: CDS
; LOCATION: (1)..(1317)
; FEATURE:
; NAME/KEY: mutation
; LOCATION: (784)..(786)
; FEATURE:
; NAME/KEY: mutation
; LOCATION: (805)..(808)
; LOCATION: (805)..(808)
US-10-157-899A-13

Query Match      100.0%; Score 41; DB 15; Length 1317;
Best Local Similarity 100.0%; Pred. No. 2.4e-08;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TTGTTGGATGCTGAGCCCCCCTACTCTATTCCGAGTATGA 41
Db      478 TTGTTGGATGCTGAGCCCCCCTACTCTATTCCGAGTATGA 518

RESULT 14
US-10-157-899A-15
; Sequence 15, Application US/10157899A
; Publication No. US20030143559A1
; GENERAL INFORMATION:
; APPLICANT: Bracken, Kathryn Rene
; APPLICANT: de los Angeles, Joseph Ernest
; APPLICANT: Huang, Ying
; APPLICANT: Kadan, Michael Joseph
; APPLICANT: Ksander, Gary Michael
; APPLICANT: Zerby, Dennis
; TITLE OF INVENTION: NOVEL ESTROGEN RECEPTOR LIGAND BINDING DOMAIN VARIANTS AND NOVEL
; TITLE OF INVENTION: LIGANDS AND PHARMACEUTICAL COMPOSITIONS
; FILE REFERENCE: 4-32018A
; CURRENT APPLICATION NUMBER: US/10/157,899A
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 60/294,839
; PRIOR FILING DATE: 2001-05-31
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 15
; LENGTH: 1317
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Fusion of a mutated human-Estrogen-receptor-ligand-binding-domain
; OTHER INFORMATION: and a zinc finger array (C7)
; NAME/KEY: CDS
; LOCATION: (1)..(1317)
; FEATURE:
; NAME/KEY: mutation
; LOCATION: (784)..(786)
; FEATURE:
; NAME/KEY: mutation
; LOCATION: (805)..(808)
; LOCATION: (805)..(808)
US-10-157-899A-15

Query Match      100.0%; Score 41; DB 15; Length 1317;
Best Local Similarity 100.0%; Pred. No. 2.4e-08;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TTGTTGGATGCTGAGCCCCCCTACTCTATTCCGAGTATGA 41
Db      478 TTGTTGGATGCTGAGCCCCCCTACTCTATTCCGAGTATGA 518
```

```
; NAME/KEY: CDS
; LOCATION: (1)..(1317)
; FEATURE:
; NAME/KEY: mutation
; LOCATION: (1093)..(1095)
US-10-157-899A-15

Query Match      100.0%; Score 41; DB 15; Length 1317;
Best Local Similarity 100.0%; Pred. No. 2.4e-08;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TTGTTGGATGCTGAGCCCCCCTACTCTATTCCGAGTATGA 41
Db      478 TTGTTGGATGCTGAGCCCCCCTACTCTATTCCGAGTATGA 518

RESULT 15
US-10-006-760-18
; Sequence 18, Application US/10006760
; Publication No. US20030186395A1
; GENERAL INFORMATION:
; APPLICANT: Koide, Shohei
; TITLE OF INVENTION: METHOD OF IDENTIFYING POLYPEPTIDE MONOBODIES WHICH BIND
; TITLE OF INVENTION: TO TARGET PROTEINS AND USE THEREOF
; FILE REFERENCE: 176/60901
; CURRENT APPLICATION NUMBER: US/10/006,760
; CURRENT FILING DATE: 2001-11-19
; PRIOR APPLICATION NUMBER: 60/249,756
; PRIOR FILING DATE: 2000-11-17
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18
; LENGTH: 1542
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: lexA-ER(alpha)EF fusion protein
US-10-006-760-18

Query Match      100.0%; Score 41; DB 15; Length 1542;
Best Local Similarity 100.0%; Pred. No. 2.5e-08;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TTGTTGGATGCTGAGCCCCCCTACTCTATTCCGAGTATGA 41
Db      703 TTGTTGGATGCTGAGCCCCCCTACTCTATTCCGAGTATGA 743

Search completed: November 3, 2004, 10:01:03
Job time : 74.6435 secs
```

This Page Blank (uspto)
BEST AVAILABLE COPY

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 3, 2004, 04:44:23 ; Search time 83.8636 Seconds
(without alignments)
2566.384 Million cell updates/sec

Title: US-09-933-267A-1_COPY_306362_306402

Perfect score: 41

Sequence: 1 ttgttgatgctgagccccc.....atactcttcgagatga 41

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 4134886 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8369772

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_23Sep04:*
1: Geneseqn1980s:*
2: Geneseqn1990s:*
3: Geneseqn2000s:*
4: Geneseqn2001as:*
5: Geneseqn2001bs:*
6: Geneseqn2002as:*
7: Geneseqn2002bs:*
8: Geneseqn2003as:*
9: Geneseqn2003bs:*
10: Geneseqn2003cs:*
11: Geneseqn2003ds:*
12: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	41	100.0	392	6	ABK89698
2	41	100.0	1044	6	ABZ23390
3	41	100.0	1044	6	ABZ23391
4	41	100.0	1218	10	ADB81478
5	41	100.0	1223	4	AAF29924
6	41	100.0	1237	6	ABK89697
7	41	100.0	1245	8	AAD53878
8	41	100.0	1245	8	AAD53879
9	41	100.0	1245	8	AAD53875
10	41	100.0	1245	8	AAD53877
11	41	100.0	1317	8	AAD53876
12	41	100.0	1317	8	AAD53882
13	41	100.0	1317	8	AAD53881
14	41	100.0	1317	8	AAD53880
15	41	100.0	1374	4	AAF29925
16	41	100.0	1380	4	AAC86919
17	41	100.0	1446	5	AAC88199
18	41	100.0	1542	10	ADP39221
19	41	100.0	1644	6	ABK89696
20	41	100.0	1770	10	ADB99351
21	41	100.0	1785	10	AAD64536

22	41	100.0	1788	6	ABL57497
23	41	100.0	1803	10	ADB93349
24	41	100.0	1956	2	AAV17756
25	41	100.0	1983	6	ABL57500
26	41	100.0	1983	6	ABL57499
27	41	100.0	1983	6	ABL57498
28	41	100.0	2028	10	AAD64590
29	41	100.0	2031	10	AAD64588
30	41	100.0	2092	1	AAW70880
31	41	100.0	2092	4	AAC86920
32	41	100.0	2092	6	ABK89700
33	41	100.0	2092	6	ABL51900
34	41	100.0	2092	8	AAD53909
35	41	100.0	2092	10	ADC09969
36	41	100.0	2106	6	ABK89699
37	41	100.0	2178	10	AAD64589
38	41	100.0	2181	10	AAD64587
39	41	100.0	2220	6	ABK56328
40	41	100.0	2220	6	ABA01104
41	41	100.0	2220	10	ADC51161
42	41	100.0	2220	10	ADC51169
43	41	100.0	2220	10	ADC51171
44	41	100.0	2220	10	ADC51170
45	41	100.0	2322	2	AAX60628

ALIGNMENTS

RESULT 1
ABK89698
ID ABK89698 standard; DNA, 392 BP.
XX
AC ABK89698;
XX
DT 05-NOV-2002 (first entry)
XX
DE Oestrogen receptor alpha nucleic acid comprising A908G mutation, #4.
XX
KW Oestrogen receptor alpha; breast cancer; pre-malignant lesion;
KW invasive breast cancer; A908G oestrogen receptor alpha transition;
KW oestrogen receptor alpha K303R substitution; human; gene; ds.
XX
OS Homo sapiens.
XX
PN WO200257283-A1.
XX
PD 25-JUL-2002.
XX
PF 16-JAN-2002; 2002WO-US004982.
XX
PR 19-JAN-2001; 2001US-0262990P.
PR 09-JUL-2001; 2001US-0304018P.
XX
PA (BAYU) BAYLOR COLLEGE MEDICINE.
XX
PI Fuqua S, O'Connell P, Allred DC, Hopp TA;
XX
DR WPI; 2002-590711/63.
XX
PT New isolated estrogen receptor alpha with A908G mutation or K303R
PT substitution, useful as diagnostic marker in breast tissue such as pre-
PT malignant lesions for the development of breast cancer, particularly
PT invasive breast cancer.
XX
PS Claim 1; Page; 133pp; English.
XX
CC The invention relates to an isolated oestrogen receptor alpha nucleic
CC acid sequence comprising an A908G mutation, or an amino acid sequence
CC comprising a K303R substitution. Also described are methods for detecting
CC susceptibility to development of breast cancer or invasive breast cancer
CC in an individual, for diagnosing breast cancer in an individual; and for
CC screening for a modulator of an oestrogen receptor alpha polypeptide

comprising a K303R substitution. The oestrogen receptor alpha is useful as a diagnostic marker in breast tissue such as pre-malignant lesions for the development of breast cancer, particularly invasive breast cancer. The methods are useful for determining susceptibility to development of breast cancer, for diagnosing, preventing or treating breast cancer. Transgenic mice may be used for screening and identifying agents that interact with the oestrogen receptor alpha, or affect breast tissue health. The A908G oestrogen receptor alpha transition is frequently present in pre-malignant lesions of the breast and can occur in the adjacent normal-appearing breast epithelium. The present sequence represents an oestrogen receptor alpha nucleic acid which may comprise the A908G mutation. Note: The present sequence is not shown in the specification but was obtained by the indexer from GenBank using the accession number given in the specification

Sequence 392 BP; 97 A; 92 C; 124 G; 79 T; 0 U; 0 Other;

Query Match 100.0%; Score 41; DB 6; Length 392;
Best Local Similarity 100.0%; Pred. No. 2.1e-07;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGTGGATGCTGAGCCCCCATACTCTATTCCGAGTATGA 41
|||||
DB 192 TTGTGGATGCTGAGCCCCCATACTCTATTCCGAGTATGA 222
|||||

RESULT 2
ABZ23390
ID ABZ23390 standard; cDNA; 1044 BP.
XX
AC ABZ23390;
XX
DT 07-APR-2003 (first entry)
XX
DE Nucleotide sequence of human oestrogen receptor alpha bait protein.
XX
KW Human; oestrogen receptor alpha; ER-alpha; CF16; CF17; CF18; CF19; CF40;
KW CF41; CF42; CF43; cofactor; osteoporosis; bone disease; reproduction;
KW hormonal dysfunction; cancer; cardiovascular disease; atherosclerosis;
KW hot flush; mood change; Alzheimer's disease; gene; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..1044
FT /tag= a
FT /product= "oestrogen receptor alpha bait"
XX
PN WO200270699-A2.
XX
PD 12-SEP-2002.
XX
PF 28-FEB-2002; 2002WO-EP002189.
XX
PR 01-MAR-2001; 2001EP-00105062.
XX
PA (LION-) LION BIOSCIENCE AG.
XX
PI Albers M, Ellwanger S, Loeser E, Koegl M;
XX
PS WPI; 2002-713451/77.
XX
PT P-PSDB; ABP70164.
XX
PT New cofactors of estrogen receptor alpha, designated as CF16, CF17, CF18, CF19, CF40, CF41, CF42 and/or CF43, useful for screening of compounds for treating osteoporosis, hormonal dysfunctions, cancer or cardiovascular diseases.
XX
PS Disclosure; Page 108-109; 111pp; English.
XX
PT The present sequence encodes a human oestrogen receptor alpha (ER-alpha) bait protein. The specification describes cofactors of ER-alpha, designated CF16, CF17, CF18, CF19, CF40, CF41, CF42, and CF43. The

cofactor polypeptides and nucleic acid molecules are useful for screening for compounds for treating osteoporosis and other bone diseases, failures in reproductive functions or hormonal dysfunctions, cancer or cardiovascular diseases such as atherosclerosis, and in preventing hot flushes, mood changes and Alzheimer's disease. The CF proteins are also useful for screening for ligands of the ER alpha. The nucleic acid sequences are useful for making vectors and CF polypeptides, transforming host cells, as research tools for developing nucleic acid probes, and for developing analytical tools such as antisense oligonucleotides

Sequence 1044 BP; 252 A; 272 C; 295 G; 225 T; 0 U; 0 Other;

Query Match 100.0%; Score 41; DB 6; Length 1044;
Best Local Similarity 100.0%; Pred. No. 2.5e-07;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGTGGATGCTGAGCCCCCATACTCTATTCCGAGTATGA 41
|||||
DB 211 TTGTGGATGCTGAGCCCCCATACTCTATTCCGAGTATGA 251
|||||

RESULT 3
ABZ23391/c
ID ABZ23391 standard; cDNA; 1044 BP.
XX
AC ABZ23391;
XX
DT 07-APR-2003 (first entry)
XX
DE Reverse complement of human oestrogen receptor alpha bait protein cDNA.
XX
KW Human; oestrogen receptor alpha; ER-alpha; CF16; CF17; CF18; CF19; CF40;
KW CF41; CF42; CF43; cofactor; osteoporosis; bone disease; reproduction;
KW hormonal dysfunction; cancer; cardiovascular disease; atherosclerosis;
KW hot flush; mood change; Alzheimer's disease; ss.
XX
OS Homo sapiens.
XX
PN WO200270699-A2.
XX
PD 12-SEP-2002.
XX
PF 28-FEB-2002; 2002WO-EP002189.
XX
PR 01-MAR-2001; 2001EP-00105062.
XX
PA (LION-) LION BIOSCIENCE AG.
XX
PI Albers M, Ellwanger S, Loeser E, Koegl M;
XX
PS WPI; 2002-713451/77.
XX
PT New cofactors of estrogen receptor alpha, designated as CF16, CF17, CF18, CF19, CF40, CF41, CF42 and/or CF43, useful for screening of compounds for treating osteoporosis, hormonal dysfunctions, cancer or cardiovascular diseases.
XX
PS Disclosure; Page 109-110; 111pp; English.
XX
PT The present sequence represents the reverse complement of an oestrogen receptor alpha (ER-alpha) bait protein. The specification describes cofactors of ER-alpha, designated CF16, CF17, CF18, CF19, CF40, CF41, CF42, and CF43. The cofactor polypeptides and nucleic acid molecules are useful for screening for compounds for treating osteoporosis and other bone diseases, failures in reproductive functions or hormonal dysfunctions, cancer or cardiovascular diseases such as atherosclerosis, and in preventing hot flushes, mood changes for ligands of the ER alpha. The nucleic acid sequences are useful for making vectors and CF polypeptides, transforming host cells, as research tools for developing nucleic acid probes, and for developing analytical tools such as antisense oligonucleotides

SQ Sequence 1044 BP; 225 A; 295 C; 272 G; 252 T; 0 U; 0 Other;

Query Match 100.0%; Score 41; DB 6; Length 1044;
Best Local Similarity 100.0%; Pred. No. 2.5e-07;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGTTGGATGCTGAGCCCCCATCTCTATTCCGAGTATGA 41
|||||
Db 834 TTGTTGGATGCTGAGCCCCCATCTCTATTCCGAGTATGA 794

RESULT 4

ADB81478
ID ADB81478 standard; DNA; 1218 BP.

XX AC ADB81478;

XX DT 04-DEC-2003 (first entry)

XX DE Human ESR-alpha with fatty acid synthase (ESR-alpha FASN) DNA.

XX KW human; ds; oestrogen receptor alpha; ESR-alpha FASN; fatty acid synthase;
KW oestrogen receptor 1; ESRI; NR3A1; bone maintenance;
KW cardiovascular system; cancer; gene therapy; hyperproliferative disease;
KW inflammation; tumour formation; infection; cytostatic; antiinflammatory;
KW antimicrobial.

XX OS Homo sapiens.

XX PN W02003052072-A2.

XX PD 26-JUN-2003.

XX PF 13-DEC-2002; 2002WO-US040083.

XX PR 18-DEC-2001; 2001US-00027983.

XX PA (ISIS-) ISIS PHARM INC.

XX PI Dobie KW, Roach MP;

XX PFPI; 2003-577322/54.

XX PT New antisense compound targeted to nucleic acid encoding estrogen
PT receptor alpha and inhibiting expression of estrogen receptor alpha,
PT useful for treating a disease or condition e.g. a hyperproliferative
PT disease.

XX PS Example 17; Page 232; 232pp; English.

XX CC This invention relates to human oestrogen receptor alpha (ESR-alpha), and
CC the novel antisense oligonucleotides that modulate its expression. The
CC oestrogen receptor alpha protein is also known as oestrogen receptor 1,
CC ESRI, and NR3A1. Oestrogen, the steroid hormone ligand of ESR-alpha, is
CC important for bone maintenance and plays a protective role in the
CC cardiovascular system, as well as being required for normal sexual
CC maturation through promoting growth and differentiation. Splice variants
CC of ESR-alpha, however, have been associated with various cancers
CC including the breast and pituitary. Accordingly, antisense
CC oligonucleotides that inhibit the expression of ESR-alpha in cells or
CC tissues can be used in gene therapy to treat conditions such as
CC hyperproliferative disease, inflammation, tumour formation and to prevent
CC or delay infection. As such, the present invention describes these
CC antisense oligos as having cytostatic, antiinflammatory and antimicrobial
CC activities. This polynucleotide is the DNA of the human oestrogen
CC receptor alpha splice variant with fatty acid synthase designated ESR-
CC alpha FASN of the invention.

XX SQ Sequence 1218 BP; 261 A; 367 C; 328 G; 262 T; 0 U; 0 Other;

Query Match

Best Local Similarity 100.0%; Score 41; DB 10; Length 1218;

Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGTTGGATGCTGAGCCCCCATCTCTATTCCGAGTATGA 41
|||||
Db 385 TTGTTGGATGCTGAGCCCCCATCTCTATTCCGAGTATGA 425

RESULT 5

AAF29924
ID AAF29924 standard; DNA; 1223 BP.

XX AC AAF29924;

XX DT 04-APR-2001 (first entry)

XX DE Human estrogen receptor alpha isoform #1 DNA.

XX KW Human; estrogen receptor alpha; cancer; osteoporosis; bone; Alzheimer's;
KW cardiovascular; ds.

XX OS Homo sapiens.

XX PN W0200100823-Al.

XX PD 04-JAN-2001.

XX PF 27-JUN-2000; 2000WO-EF005981.

XX PR 29-JUN-1999; 99IT-MT001433.

XX PA (EUMO-) EURO MOLECULAR BIOLOGY LAB.

XX PI Gannon F, Dengler S, Flouriot G;

XX PFPI; 2001-137955/14.

XX PT Novel isoforms of human estrogen receptor alpha useful for preparing
PT therapeutic agents for treating cancer, osteoporosis, Alzheimer's disease
PT and cardiovascular diseases.

XX PS Claim 1; Pag e44; 53pp; English.

XX CC The present invention relates to a human estrogen receptor (hER)-alpha
CC isoform. Molecules which modulate the activity of the estrogen receptor
CC are useful for the preparation of therapeutic agents for treating cancer,
CC osteoporosis and other bone disorders, Alzheimer's disease and
CC cardiovascular diseases

XX SQ Sequence 1223 BP; 303 A; 316 C; 341 G; 263 T; 0 U; 0 Other;

Query Match 100.0%; Score 41; DB 4; Length 1223;

Best Local Similarity 100.0%; Pred. No. 2.5e-07;

Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGTTGGATGCTGAGCCCCCATCTCTATTCCGAGTATGA 41
|||||

Db 390 TTGTTGGATGCTGAGCCCCCATCTCTATTCCGAGTATGA 430

RESULT 6

ABK89697
ID ABK89697 standard; DNA; 1237 BP.

XX AC ABK89697;

XX DT 05-NOV-2002 (first entry)

XX DE Oestrogen receptor alpha nucleic acid comprising A908G mutation, #3.

XX KW Oestrogen receptor alpha; breast cancer; pre-malignant lesion;
KW invasive breast cancer; A908G oestrogen receptor alpha transition;

XX KW oestrogen receptor alpha K303R substitution; human; gene; ds.

XX OS Homo sapiens.

```
XX PN WO200257283-A1.
XX PD
XX PF
XX PF 25-JUL-2002.
XX PF 16-JAN-2002; 2002WO-US004982.
XX PR 19-JAN-2001; 2001US-0262990P.
XX PR 09-JUL-2001; 2001US-0304018P.
XX PA (BAYU ) BAYLOR COLLEGE MEDICINE.
XX PI Fuqua S, O'connell P, Allred DC, Hopp TA;
XX DR WPI; 2002-590711/63.
XX PR New isolated estrogen receptor alpha with A908G mutation or K303R
XX PT substitution, useful as diagnostic marker in breast tissue such as pre-
XX PT malignant lesions for the development of breast cancer, particularly
XX PT invasive breast cancer.
XX PS Claim 1; Page; 133pp; English.
XX CC The invention relates to an isolated oestrogen receptor alpha nucleic
XX CC acid sequence comprising an A908G mutation, or an amino acid sequence
XX CC comprising a K303R substitution. Also described are methods for detecting
XX CC susceptibility to development of breast cancer or invasive breast cancer
XX CC in an individual, for diagnosing breast cancer in an individual; and for
XX CC screening for a modulator of an oestrogen receptor alpha polypeptide
XX CC comprising a K303R substitution. The oestrogen receptor alpha is useful
XX CC as a diagnostic marker in breast tissue such as pre-malignant lesions
XX CC for the development of breast cancer, particularly invasive breast
XX CC cancer. The methods are useful for determining susceptibility to
XX CC development of breast cancer, for diagnosing, preventing or treating
XX CC breast cancer. Transgenic mice may be used for screening and identifying
XX CC agents that interact with the oestrogen receptor alpha, or affect breast
XX CC tissue health. The A908G oestrogen receptor alpha transition is
XX CC frequently present in pre-malignant lesions of the breast and can occur
XX CC in the adjacent normal-appearing breast epithelium. The present sequence
XX CC represents an oestrogen receptor alpha nucleic acid which may comprise
XX CC the A908G mutation. Note: The present sequence is not shown in the
XX CC specification but was obtained by the indexer from GenBank using the
XX CC accession number given in the specification
XX SQ Sequence 1237 BP; 369 A; 218 C; 286 G; 364 T; 0 U; 0 Other;
XX
XX Query Match 100.0%; Score 41; DB 6; Length 1237;
XX Best Local Similarity 100.0%; Pred. No. 2.5e-07;
XX Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 TTGTTGGATGCTGAGCCCCCCTACTCTATTCCGAGTATGA 41
XX DB 544 TTGTTGGATGCTGAGCCCCCCTACTCTATTCCGAGTATGA 584
XX
XX RESULT 7
XX AAD53878
XX ID AAD53878 standard; DNA; 1245 BP.
XX AC AAD53878;
XX XX
XX DT 28-MAY-2003 (first entry)
XX DE C7LBDAS chimeric mutant DNA #2.
XX XX
XX XX Oestrogen receptor; ER alpha; ligand binding domain; genetic disease;
XX KW acquired disease; cell proliferative disorder; cancer; adenocarcinoma;
XX KW LBD; gene switch; transgenic animal; transgenic; gene therapy; human;
XX KW chimeric; zinc finger array; C7; mutant; gene; ds.
XX OS Homo sapiens.
XX OS Unidentified.
XX OS Chimeric.
```

```
XX Key Location/Qualifiers
XX CDS 1..1245
XX FT /*tag= a
XX FT /product= "C7LBDAS fusion protein"
XX FT mutation replace(712,A)
XX FT /*tag= b
XX FT mutation replace(733..735, CTG)
XX FT /*tag= c
XX
XX WO200297050-A2.
XX
XX 05-DEC-2002.
XX
XX 31-MAY-2002; 2002WO-US016946.
XX
XX 31-MAY-2001; 2001US-0294839P.
XX
XX (NOVS ) NOVARTIS AG.
XX
XX Bracken KS, De Los Angeles JE, Huang Y, Kadan MJ, Keander GW;
XX PI Zerby DB;
XX
XX WPI; 2003-156794/15.
XX P-PSDB; AAE35278.
XX
XX New mutant estrogen receptor ligand binding domain capable of interacting
XX with non-endogenous ligand, useful e.g. in combination with a ligand for
XX constructing selective molecular gene switches for regulating gene
XX function.
XX
XX Example 4; Page 126-127; 159pp; English.
XX
XX The invention relates to a mutant oestrogen receptor (ER) alpha-ligand
XX binding domain (LBD) which comprises an amino acid modification in region
XX 1, region 2 or both and interacting with a non-endogenous ligand as a
XX result of the amino acid modification. Sequences of the invention are
XX useful for treatment of genetic diseases, acquired diseases and any other
XX conditions including cell proliferative disorders such as cancer e.g.
XX lung, breast, lymphoid, gastrointestinal, genito-urinary tract
XX adenocarcinomas and other malignancies such as colon cancers, renal-cell
XX carcinoma, prostate cancer, non-small cell carcinoma of the lung, cancer
XX of the small intestine and cancer of the oesophagus. The invention is
XX useful for constructing selective molecular gene switches for regulating
XX gene function in plants and transgenic animals. It is also useful in gene
XX therapy. The present sequence is human ER alpha LBD -zinc finger array
XX (C7) chimeric mutant DNA
XX
XX SQ Sequence 1245 BP; 281 A; 365 C; 328 G; 271 T; 0 U; 0 Other;
XX
XX Query Match 100.0%; Score 41; DB 8; Length 1245;
XX Best Local Similarity 100.0%; Pred. No. 2.5e-07;
XX Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 TTGTTGGATGCTGAGCCCCCCTACTCTATTCCGAGTATGA 41
XX DB 406 TTGTTGGATGCTGAGCCCCCCTACTCTATTCCGAGTATGA 446
XX
XX RESULT 8
XX AAD53879
XX ID AAD53879 standard; DNA; 1245 BP.
XX XX
XX AC AAD53879;
XX XX
XX DT 28-MAY-2003 (first entry)
XX DE C7LBDAS chimeric mutant DNA #3.
XX XX
XX XX Oestrogen receptor; ER alpha; ligand binding domain; genetic disease;
XX KW acquired disease; cell proliferative disorder; cancer; adenocarcinoma;
XX KW LBD; gene switch; transgenic animal; transgenic; gene therapy; human;
XX KW chimeric; zinc finger array; C7; mutant; gene; ds.
```



```
XX OS Homo sapiens.
XX OS Unidentified.
XX OS Chimeric.
XX FH Key Location/Qualifiers
XX FT CDS 1..1245
XX FT /*tag= a
XX FT /product= "C7LBDAS fusion protein"
XX FT mutation replace(1021..1023, CAT)
XX FT /*tag= b
XX PN WO200297050-A2.
XX XX
XX PD 05-DEC-2002.
XX XX
XX PF 31-MAY-2002; 2002WO-US016946.
XX XX
XX PR 31-MAY-2001; 2001US-0294839P.
XX XX
XX PA (NOVS ) NOVARTIS AG.
XX XX
XX PI Bracken KR, De Los Angeles JE, Huang Y, Kadan MJ, Ksander GM;
XX PI Zerby DB;
XX XX
XX DR WPI; 2003-156794/15.
XX DR P-PSDB; AAE35279.
XX XX
XX PT New mutant estrogen receptor ligand binding domain capable of interacting
XX PT with non-endogenous ligand, useful e.g. in combination with a ligand for
XX PT constructing selective molecular gene switches for regulating gene
XX PT function.
XX PS Example 4; Page 129-131; 159pp; English.
XX CC The invention relates to a mutant oestrogen receptor (ER) alpha-ligand
XX CC binding domain (LBD) which comprises an amino acid modification in region
XX CC 1, region 2 or both and interacting with a non-endogenous ligand as a
XX CC result of the amino acid modification. Sequences of the invention are
XX CC useful for treatment of genetic diseases, acquired diseases and any other
XX CC conditions including cell proliferative disorders such as cancer e.g.
XX CC lung, breast, lymphoid, gastrointestinal, genito-urinary tract
XX CC adenocarcinomas and other malignancies such as colon cancers, renal- cell
XX CC carcinoma, prostate cancer, non-small cell carcinoma of the lung, cancer
XX CC of the small intestine and cancer of the oesophagus. The invention is
XX CC useful for constructing selective molecular gene switches for regulating
XX CC gene function in plants and transgenic animals. It is also useful in gene
XX CC therapy. The present sequence is human ER alpha LBD -zinc finger array
XX CC (C7) chimeric mutant DNA
XX SQ Sequence 1245 BP; 281 A; 364 C; 329 G; 271 T; 0 U; 0 Other;
XX
XX Query Match 100.0%; Score 41; DB 8; Length 1245;
XX Best Local Similarity 100.0%; Pred. No. 2.5e-07;
XX Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 TTGTTGGATGCTGAGCCCCCATCTACTCTATTCGAGTATGA 41
XX DB 406 TTGTTGGATGCTGAGCCCCCATCTACTCTATTCGAGTATGA 446
XX
XX RESULT 9
XX AAD53875
XX ID AAD53875 standard; DNA; 1245 BP.
XX AC AAD53875;
XX XX
XX DT 28-MAY-2003 (first entry)
XX XX
XX DE C7LBDAS chimeric DNA #1.
XX XX
XX KW Oestrogen receptor; ER alpha; ligand binding domain; genetic disease;
XX KW acquired disease; cell proliferative disorder; cancer; adenocarcinoma;
```

```
LBD; gene switch; transgenic animal; transgenic; gene therapy; human;
chimeric; zinc finger array; C7; gene; ds.
Homo sapiens.
Unidentified.
Chimeric.
Key Location/Qualifiers
CDS 1..1245
/*tag= a
/product= "C7LBDAS fusion protein"
WO200297050-A2.
05-DEC-2002.
31-MAY-2002; 2002WO-US016946.
31-MAY-2001; 2001US-0294839P.
(NOVS ) NOVARTIS AG.
Bracken KR, De Los Angeles JE, Huang Y, Kadan MJ, Ksander GM;
Zerby DB;
WPI; 2003-156794/15.
P-PSDB; AAE35275.
New mutant estrogen receptor ligand binding domain capable of interacting
with non-endogenous ligand, useful e.g. in combination with a ligand for
constructing selective molecular gene switches for regulating gene
function.
Example 4; Page 115-117; 159pp; English.
The invention relates to a mutant oestrogen receptor (ER) alpha-ligand
binding domain (LBD) which comprises an amino acid modification in region
1, region 2 or both and interacting with a non-endogenous ligand as a
result of the amino acid modification. Sequences of the invention are
useful for treatment of genetic diseases, acquired diseases and any other
conditions including cell proliferative disorders such as cancer e.g.
lung, breast, lymphoid, gastrointestinal, genito-urinary tract
adenocarcinomas and other malignancies such as colon cancers, renal- cell
carcinoma, prostate cancer, non-small cell carcinoma of the lung, cancer
of the small intestine and cancer of the oesophagus. The invention is
useful for constructing selective molecular gene switches for regulating
gene function in plants and transgenic animals. It is also useful in gene
therapy. The present sequence is human ER alpha LBD -zinc finger array
(C7) chimeric DNA
Sequence 1245 BP; 282 A; 364 C; 327 G; 272 T; 0 U; 0 Other;
Query Match 100.0%; Score 41; DB 8; Length 1245;
Best Local Similarity 100.0%; Pred. No. 2.5e-07;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TTGTTGGATGCTGAGCCCCCATCTACTCTATTCGAGTATGA 41
DB 406 TTGTTGGATGCTGAGCCCCCATCTACTCTATTCGAGTATGA 446
RESULT 10
AAD53877
ID AAD53877 standard; DNA; 1245 BP.
XX AC AAD53877;
XX XX
XX DT 28-MAY-2003 (first entry)
XX XX
XX DE C7LBDAS chimeric mutant DNA #1.
XX KW Oestrogen receptor; ER alpha; ligand binding domain; genetic disease;
XX KW acquired disease; cell proliferative disorder; cancer; adenocarcinoma;
```

W	LBD; gene switch; transgenic animal; transgenic; gene therapy; human;	DT	28-MAY-2003 (first entry)
W	chimeric; zinc finger array; C7; mutant; gene; ds.	DE	C7LBDBS chimeric DNA #1.
XX		XX	Oestrogen receptor; ER alpha; ligand binding domain; genetic disease;
OS	Homo sapiens.	KW	acquired disease; cell proliferative disorder; cancer; adenocarcinoma;
OS	Unidentified.	KW	LBD; gene switch; transgenic animal; transgenic; gene therapy; human;
OS	Chimeric.	KW	chimeric; zinc finger array; C7; gene; ds.
XX		XX	Homo sapiens.
FT	Key	OS	Unidentified.
FT	CDS	OS	Chimeric.
FT	Location/Qualifiers	XX	Key
FT	1..1245	XX	Location/Qualifiers
FT	/*tag= a	FT	1..1317
FT	replace(613,A)	FT	/*tag= a
FT	/*tag= b	FT	/product= "C7LBDBS fusion protein"
FT	replace(721..722, AT)	XX	
FT	/*tag= c	PN	WO200297050-A2.
FT	replace(733..735, CTG)	XX	
FT	/*tag= d	XX	
XX	WO200297050-A2.	PD	05-DEC-2002.
XX		XX	
XX		XX	
XX		PF	31-MAY-2002; 2002WO-US016946.
XX		XX	
XX		XX	31-MAY-2001; 2001US-0294839P.
XX		XX	
XX		XX	(NOVS) NOVARTIS AG.
XX		PA	
XX		XX	
XX		PI	Bracken KR, De Los Angeles JE, Huang Y, Kadan MJ, Ksander GM;
XX		PI	Zerby DB;
XX		XX	
XX		XX	
XX		DR	WPI; 2003-156794/15.
XX		DR	P-PSDB; AAE35276.
XX		XX	
XX		PT	New mutant estrogen receptor ligand binding domain capable of interacting
XX		PT	with non-endogenous ligand, useful e.g. in combination with a ligand for
XX		PT	constructing selective molecular gene switches for regulating gene
XX		PT	function.
XX		XX	
XX		PS	Example 4; Page 118-120; 159pp; English.
XX		XX	
XX		CC	The invention relates to a mutant oestrogen receptor (ER) alpha-ligand
XX		CC	binding domain (LBD) which comprises an amino acid modification in region
XX		CC	1, region 2 or both and interacting with a non-endogenous ligand as a
XX		CC	result of the amino acid modification. Sequences of the invention are
XX		CC	useful for treatment of genetic diseases, acquired diseases and any other
XX		CC	conditions including cell proliferative disorders such as cancer e.g.
XX		CC	lung, breast, lymphoid, gastrointestinal, genito-urinary tract
XX		CC	adenocarcinomas and other malignancies such as colon cancers, renal- cell
XX		CC	carcinoma, prostate cancer, non-small cell carcinoma of the lung, cancer
XX		CC	of the small intestine and cancer of the oesophagus. The invention is
XX		CC	useful for constructing selective molecular gene switches for regulating
XX		CC	gene function in plants and transgenic animals. It is also useful in gene
XX		CC	therapy. The present sequence is human ER alpha LBD -zinc finger array
XX		CC	(C7) chimeric mutant DNA
XX		XX	
XX		SQ	Sequence 1317 BP; 305 A; 372 C; 361 G; 279 T; 0 U; 0 Other;
XX			
XX			Query Match 100.0%; Score 41; DB 8; Length 1317;
XX			Best Local Similarity 100.0%; Pred. No. 2.6e-07;
XX			Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX			
XX		QY	1 TTGTTGGATGCTGAGCCCCCATCTCTATTCCGAGTATGA 41
XX			
XX		DB	478 TTGTTGGATGCTGAGCCCCCATCTCTATTCCGAGTATGA 518
XX			
XX			RESULT 12
XX			AAD53882
XX		ID	AAD53882 standard; DNA; 1317 BP.
XX		XX	
XX		AC	AAD53882;
XX		XX	

DT 28-MAY-2003 (first entry)
XX C7LBDBS chimeric mutant DNA #3.
DE
XX Oestrogen receptor; ER alpha; ligand binding domain; genetic disease;
KW acquired disease; cell proliferative disorder; cancer; adenocarcinoma;
KW LBD; gene switch; transgenic animal; transgenic; gene therapy; human;
KW chimeric; zinc finger array; C7; mutant; gene; ds.
XX
OS Homo sapiens.
OS Unidentified.
OS Chimeric.
XX
XX Key Location/Qualifiers
XX CDS 1..1317
XX /*tag= a
XX /product= "C7LBDBS fusion protein"
XX mutation replace(1093..1095, CAT)
XX /*tag= b
XX
XX WO200297050-A2.
XX
XX 05-DEC-2002.
XX
XX 31-MAY-2002; 2002WO-US016946.
XX
XX 31-MAY-2001; 2001US-0294839P.
XX
XX (NOVS) NOVARTIS AG.
XX
XX Bracken KR, De Los Angeles JE, Huang Y, Kadan MJ, Ksander GM;
XX Zerby DB;
XX
XX WPI; 2003-156794/15.
XX P-PSDB; AAE35282.
XX
XX New mutant estrogen receptor ligand binding domain capable of interacting
XX with non-endogenous ligand, useful e.g. in combination with a ligand for
XX constructing selective molecular gene switches for regulating gene
XX function.
XX
XX Example 4; Page 140-142; 159pp; English.
XX
XX The invention relates to a mutant oestrogen receptor (ER) alpha-ligand
XX binding domain (LBD) which comprises an amino acid modification in region
XX 1, region 2 or both and interacting with a non-endogenous ligand as a
XX result of the amino acid modification. Sequences of the invention are
XX useful for treatment of genetic diseases, acquired diseases and any other
XX conditions including cell proliferative disorders such as cancer e.g.
XX lung, breast, lymphoid, gastrointestinal, genito-urinary tract
XX adenocarcinomas and other malignancies such as colon cancers, renal- cell
XX carcinoma, prostate cancer, non-small cell carcinoma of the lung, cancer
XX of the small intestine and cancer of the oesophagus. The invention is
XX useful for constructing selective molecular gene switches for regulating
XX gene function in plants and transgenic animals. It is also useful in gene
XX therapy. The present sequence is human ER alpha LBD -zinc finger array
XX (C7) chimeric mutant DNA
XX
SQ Sequence 1317 BP; 304 A; 372 C; 363 G; 278 T; 0 U; 0 Other;
Query Match 100.0%; Score 41; DB 8; Length 1317;
Best Local Similarity 100.0%; Pred. No. 2.6e-07;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TTGTTGGATGCTGAGCCCCCATCTACTCTATTCGAGTATGA 41
DB 478 TTGTTGGATGCTGAGCCCCCATCTACTCTATTCGAGTATGA 518
RESULT 13
AAD53881
ID AAD53881 standard; DNA; 1317 BP.
XX

AC AAD53881;
XX
XX 28-MAY-2003 (first entry)
XX
XX C7LBDBS chimeric mutant DNA #2.
DE
XX Oestrogen receptor; ER alpha; ligand binding domain; genetic disease;
KW acquired disease; cell proliferative disorder; cancer; adenocarcinoma;
KW LBD; gene switch; transgenic animal; transgenic; gene therapy; human;
KW chimeric; zinc finger array; C7; mutant; gene; ds.
XX
XX Homo sapiens.
XX Unidentified.
XX Chimeric.
XX
XX Key Location/Qualifiers
XX CDS 1..1317
XX /*tag= a
XX /product= "C7LBDBS fusion protein"
XX mutation replace(784,A)
XX /*tag= b
XX mutation replace(804..807, CTG)
XX /*tag= c
XX
XX WO200297050-A2.
XX
XX 05-DEC-2002.
XX
XX 31-MAY-2002; 2002WO-US016946.
XX
XX 31-MAY-2001; 2001US-0294839P.
XX
XX (NOVS) NOVARTIS AG.
XX
XX Bracken KR, De Los Angeles JE, Huang Y, Kadan MJ, Ksander GM;
XX Zerby DB;
XX
XX WPI; 2003-156794/15.
XX P-PSDB; AAE35281.
XX
XX New mutant estrogen receptor ligand binding domain capable of interacting
XX with non-endogenous ligand, useful e.g. in combination with a ligand for
XX constructing selective molecular gene switches for regulating gene
XX function.
XX
XX Example 4; Page 136-138; 159pp; English.
XX
XX The invention relates to a mutant oestrogen receptor (ER) alpha-ligand
XX binding domain (LBD) which comprises an amino acid modification in region
XX 1, region 2 or both and interacting with a non-endogenous ligand as a
XX result of the amino acid modification. Sequences of the invention are
XX useful for treatment of genetic diseases, acquired diseases and any other
XX conditions including cell proliferative disorders such as cancer e.g.
XX lung, breast, lymphoid, gastrointestinal, genito-urinary tract
XX adenocarcinomas and other malignancies such as colon cancers, renal- cell
XX carcinoma, prostate cancer, non-small cell carcinoma of the lung, cancer
XX of the small intestine and cancer of the oesophagus. The invention is
XX useful for constructing selective molecular gene switches for regulating
XX gene function in plants and transgenic animals. It is also useful in gene
XX therapy. The present sequence is human ER alpha LBD -zinc finger array
XX (C7) chimeric mutant DNA
XX
SQ Sequence 1317 BP; 304 A; 373 C; 362 G; 278 T; 0 U; 0 Other;
Query Match 100.0%; Score 41; DB 8; Length 1317;
Best Local Similarity 100.0%; Pred. No. 2.6e-07;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TTGTTGGATGCTGAGCCCCCATCTACTCTATTCGAGTATGA 41
DB 478 TTGTTGGATGCTGAGCCCCCATCTACTCTATTCGAGTATGA 518

```
RESULT 14
AAD53880
ID AAD53880 standard; DNA; 1317 BP.
XX
AC AAD53880;
XX
XX 28-MAY-2003 (first entry)
XX
XX C7LBDBS chimeric mutant DNA #1.
XX
KW Oestrogen receptor; ER alpha; ligand binding domain; genetic disease;
KW acquired disease; cell proliferative disorder; cancer; adenocarcinoma;
KW LBD; gene switch; transgenic animal; transgenic; gene therapy; human;
KW chimeric; zinc finger array; C7; mutant; gene; ds.
XX
XX Homo sapiens.
OS
OS Unidentified.
OS Chimeric.
XX
XX Key Location/Qualifiers
XX CDS 1..1317
XX FT /*tag= a
XX FT /product= "C7LBDBS fusion protein"
XX FT replace(685,A)
XX FT /*tag= b
XX FT mutation
XX FT replace(793..794, AT)
XX FT /*tag= c
XX FT mutation
XX FT replace(805..807, CTG)
XX FT /*tag= d
XX
XX WO200297050-A2.
XX
XX 05-DEC-2002.
XX
XX 31-MAY-2002; 2002WO-US016946.
XX
XX 31-MAY-2001; 2001US-0294839P.
XX
XX (NOVS ) NOVARTIS AG.
XX
XX Bracken KR, De Los Angeles JE, Huang Y, Kadan MJ, Keander GM;
XX Zerby DB;
XX
XX WPI; 2003-156794/15.
XX P-PSDB; AAE35280.
XX
XX New mutant estrogen receptor ligand binding domain capable of interacting
XX with non-endogenous ligand, useful e.g. in combination with a ligand for
XX constructing selective molecular gene switches for regulating gene
XX function.
XX
XX Example 4; Page 133-134; 159pp; English.
XX
XX The invention relates to a mutant oestrogen receptor (ER) alpha-ligand
XX binding domain (LBD) which comprises an amino acid modification in region
XX 1, region 2 or both and interacting with a non-endogenous ligand as a
XX result of the amino acid modification. Sequences of the invention are
XX useful for treatment of genetic diseases, acquired diseases and any other
XX conditions including cell proliferative disorders such as cancer e.g.
XX lung, breast, lymphoid, gastrointestinal, genito-urinary tract
XX adenocarcinomas and other malignancies such as colon cancers, renal- cell
XX carcinoma, prostate cancer, non-small cell carcinoma of the lung, cancer
XX of the small intestine and cancer of the oesophagus. The invention is
XX useful for constructing selective molecular gene switches for regulating
XX gene function in plants and transgenic animals. It is also useful in gene
XX therapy. The present sequence is human ER alpha LBD -zinc finger array
XX (C7) chimeric mutant DNA
XX
XX Sequence 1317 BP; 304 A; 373 C; 362 G; 278 T; 0 U; 0 Other;
Query Match 100.0%; Score 41; DB 8; Length 1317;
Best Local Similarity 100.0%; Pred. No. 2.6e-07;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 TTGTTGGATGCTGAGCCCCCATACTCTATTCCGAGTATGA 41
|||
Db 478 TTGTTGGATGCTGAGCCCCCATACTCTATTCCGAGTATGA 518
```

RESULT 15

```
AAF29925
ID AAF29925 standard; DNA; 1374 BP.
XX
AC AAF29925;
XX
DT 04-APR-2001 (first entry)
XX
DE Human estrogen receptor alpha isoform #2 DNA.
XX
XX Human; estrogen receptor alpha; cancer; osteoporosis; bone; Alzheimer's;
KW cardiovascular; ds.
XX
XX Homo sapiens.
OS
XX WO200100823-A1.
XX
XX 04-JAN-2001.
XX
XX 27-JUN-2000; 2000WO-EP005981.
XX
XX 29-JUN-1999; 99IT-MI001433.
XX
XX (EUMO-) EURO MOLECULAR BIOLOGY LAB.
XX
XX Gannon F, Denger S, Flouriot G;
XX WPI; 2001-137955/14.
XX
XX Novel isoforms of human estrogen receptor alpha useful for preparing
XX therapeutic agents for treating cancer, osteoporosis, Alzheimer's disease
XX and cardiovascular diseases.
XX
XX Claim 1; Page 44-45; 53pp; English.
XX
XX The present invention relates to a human estrogen receptor (hER)-alpha
XX isoform. Molecules which modulate the activity of the estrogen receptor
XX are useful for the preparation of therapeutic agents for treating cancer,
XX osteoporosis and other bone disorders, Alzheimer's disease and
XX cardiovascular diseases
XX
XX Sequence 1374 BP; 349 A; 346 C; 386 G; 293 T; 0 U; 0 Other;
Query Match 100.0%; Score 41; DB 4; Length 1374;
Best Local Similarity 100.0%; Pred. No. 2.6e-07;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 TTGTTGGATGCTGAGCCCCCATACTCTATTCCGAGTATGA 41
|||
Db 541 TTGTTGGATGCTGAGCCCCCATACTCTATTCCGAGTATGA 581
```

Search completed: November 3, 2004, 06:59:25
Job time : 84.8636 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 3, 2004, 05:58:29 ; Search time 18.8325 Seconds
(without alignments)
1547.447 Million cell updates/sec

Title: US-09-933-267A-1_COPY_306362_306402

Perfect score: 41

Sequence: 1 ttgttgatgctgagcccc.....atactattccgagatga 41

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 824507 seqs, 355394441 residues

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents NA:*

- 1: /cgn2_6/ptodata/1/ina/5A-COMB.seq:*
- 2: /cgn2_6/ptodata/1/ina/5B-COMB.seq:*
- 3: /cgn2_6/ptodata/1/ina/6A-COMB.seq:*
- 4: /cgn2_6/ptodata/1/ina/6B-COMB.seq:*
- 5: /cgn2_6/ptodata/1/ina/PTUS-COMB.seq:*
- 6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	41	100.0	1218	4	US-10-027-983-98
2	41	100.0	1956	3	US-08-693-940-2
3	41	100.0	1956	4	US-09-566-660-2
4	41	100.0	2322	3	US-08-564-264-2
5	41	100.0	4963	1	US-08-076-726-16
6	41	100.0	4963	1	US-08-260-452-9
7	41	100.0	4963	2	US-08-481-970-9
8	41	100.0	4963	2	US-08-897-719-9
9	41	100.0	4963	3	US-09-163-269-9
10	41	100.0	4963	4	US-09-281-674-9
11	41	100.0	6450	3	US-09-041-886-34
12	41	100.0	6450	4	US-08-453-998-1
13	41	100.0	6450	4	US-10-027-983-3
14	41	100.0	8566	4	US-10-027-983-10
15	21.6	52.7	334	4	US-09-270-767-8774
16	21.6	52.7	334	4	US-09-270-767-24056
17	21	51.2	976	3	US-08-361-083-79
18	21	51.2	976	4	US-09-536-784-79
19	21	51.2	1929	4	US-09-583-110-1148
20	21	51.2	11384	4	US-08-961-527-45
21	20.4	49.8	3182	4	US-09-221-017B-1029
22	20	48.8	1230025	4	US-09-198-452A-1
23	19.8	48.3	9458	4	US-08-356-171B-114
24	19.8	48.3	9458	4	US-08-781-986A-114
25	19.8	48.3	17041	1	US-08-076-011-1
26	19.4	47.3	432	4	US-09-513-999C-2913
27	19.4	47.3	202001	4	US-09-734-674-3

C 28	19.2	46.8	1134	4	US-09-270-767-11368	Sequence 11368, A
C 29	19.2	46.8	35081	2	US-08-752-760A-1	Sequence 1, Appli
C 30	19	46.3	1000	4	US-09-641-638-460	Sequence 460, App
C 31	19	46.3	1000	4	US-10-170-037-460	Sequence 460, App
C 32	19	46.3	1809	4	US-09-489-039A-36	Sequence 36, Appli
C 33	19	46.3	4266	4	US-09-489-039A-7	Sequence 7, Appli
C 34	18.8	45.9	252	4	US-09-313-294A-2329	Sequence 2329, Ap
C 35	18.8	45.9	446	4	US-09-513-999C-1069	Sequence 1069, Ap
C 36	18.8	45.9	1104	4	US-09-328-352-3352	Sequence 3352, Ap
C 37	18.8	45.9	319608	4	US-09-539-333D-1	Sequence 1, Appli
C 38	18.8	45.9	319608	4	US-09-679-409-1	Sequence 1, Appli
C 39	18.6	45.4	271	4	US-09-513-999C-8712	Sequence 8712, Ap
C 40	18.6	45.4	1242	3	US-09-068-132-1	Sequence 1, Appli
C 41	18.6	45.4	1242	3	US-09-068-132-2	Sequence 2, Appli
C 42	18.6	45.4	1242	4	US-10-018-730A-1	Sequence 1, Appli
C 43	18.6	45.4	3501	4	US-09-643-593-4	Sequence 4, Appli
C 44	18.6	45.4	9897	4	US-08-961-527-10	Sequence 10, Appli
C 45	18.6	45.4	17425	3	US-09-511-625B-5	Sequence 5, Appli

ALIGNMENTS

RESULT 1

US-10-027-983-98

; Sequence 98, Application US/10027983

; Patent No. 6617162

; GENERAL INFORMATION:

; APPLICANT: Kenneth W. Dobie

; APPLICANT: Mark P. Roach

; TITLE OF INVENTION: ANTISENSE MODULATION OF ESTROGEN RECEPTOR ALPHA EXPRESSION

; FILE REFERENCE: RTS-0340

; CURRENT APPLICATION NUMBER: US/10/027,983

; CURRENT FILING DATE: 2001-12-18

; NUMBER OF SEQ ID NOS: 98

; SEQ ID NO 98

; LENGTH: 1218

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

US-10-027-983-98

Query Match 100.0%; Score 41; DB 4; Length 1218;

Best Local Similarity 100.0%; Pred. No. 9.6e-09;

Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGTTGGATGCTGAGCCCCCATCTCTATTCCGAGTATGA 41

Db 385 TTGTTGGATGCTGAGCCCCCATCTCTATTCCGAGTATGA 425

RESULT 2

US-08-693-940-2

; Sequence 2, Application US/08693940

; Patent No. 6133027

; GENERAL INFORMATION:

; APPLICANT: Yee, Jiling-Kuan

; APPLICANT: Friedman, Theodore

; APPLICANT: Chen, Shan-Tai

; TITLE OF INVENTION: Inducible Expression System

; TITLE OF INVENTION: Useful in the Generation of Packaging Cell Lines for

; TITLE OF INVENTION: Pseudotyped Retroviral Vectors

; NUMBER OF SEQUENCES: 3

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Bozicevic & Reed, LLP

; STREET: 285 Hamilton Ave, Suite 200

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

; ZIP: 94301

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

```
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/593,940
; FILING DATE: 07-AUG-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Francis, Carol L
; REGISTRATION NUMBER: 36,513
; REFERENCE/DOCKET NUMBER: 6510-055001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-327-3400
; TELEFAX: 650-327-3231
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1956 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 1...1953
; OTHER INFORMATION:
;
US-08-593-940-2

Query Match 100.0%; Score 41; DB 3; Length 1956;
Best Local Similarity 100.0%; Pred. No. 1.1e-08;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGTTGATGCTGAGCCCCCATACTCTATTCCGAGTATGA 41
Db 1123 TTGTTGATGCTGAGCCCCCATACTCTATTCCGAGTATGA 1163

RESULT 3
US-09-566-660-2
; Sequence 2, Application US/09566660
; Patent No. 6432705
; GENERAL INFORMATION:
; APPLICANT: Yee, Jiling-Kuan
; APPLICANT: Friedmann, Theodore
; APPLICANT: Chen, Shin-Tai
; TITLE OF INVENTION: Inducible Expression System
; FILE REFERENCE: 6510-055001
; CURRENT APPLICATION NUMBER: US/09/566,660
; CURRENT FILING DATE: 2000-05-08
; PRIOR APPLICATION NUMBER: 08/693,940
; PRIOR FILING DATE: 1996-08-07
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1956
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1953)
; OTHER INFORMATION: Multi-chimeric transactivating factor
;
US-09-566-660-2

Query Match 100.0%; Score 41; DB 4; Length 1956;
Best Local Similarity 100.0%; Pred. No. 1.1e-08;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGTTGATGCTGAGCCCCCATACTCTATTCCGAGTATGA 41
Db 1123 TTGTTGATGCTGAGCCCCCATACTCTATTCCGAGTATGA 1163

; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/593,940
; FILING DATE: 07-AUG-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Francis, Carol L
; REGISTRATION NUMBER: 36,513
; REFERENCE/DOCKET NUMBER: 6510-055001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-327-3400
; TELEFAX: 650-327-3231
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1956 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 1...1953
; OTHER INFORMATION:
;
US-08-593-940-2

Query Match 100.0%; Score 41; DB 3; Length 1956;
Best Local Similarity 100.0%; Pred. No. 1.1e-08;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGTTGATGCTGAGCCCCCATACTCTATTCCGAGTATGA 41
Db 1123 TTGTTGATGCTGAGCCCCCATACTCTATTCCGAGTATGA 1163

RESULT 4
US-08-564-264-2
; Sequence 2, Application US/08564264
; Patent No. 6040430
; GENERAL INFORMATION:
; APPLICANT: STEWART, Francis
; TITLE OF INVENTION: REGULATION OF SITE-SPECIFIC
; RECOMBINATION BY SITE-SPECIFIC RECOMBINASE/NUCLEAR
; TITLE OF INVENTION: RECEPTOR FUSION PROTEINS
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nikaido, Marmelstein, Murray & Oram
; STREET: 655 Fifteenth Street N.W. Suite 330
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-5701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/564,264
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP94/02088
; FILING DATE: 28-JUN-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 93 110 298.2
; FILING DATE: 28-JUN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Murray, Robert B.
; REGISTRATION NUMBER: 22,980
; REFERENCE/DOCKET NUMBER: P564-5019
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)638-5000
; TELEFAX: (202)638-4810
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2322 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1-1269)
; OTHER INFORMATION: /note= "FLP recombinase domain."
;
; NAME/KEY: misc feature
; LOCATION: (1270-1284)
; OTHER INFORMATION: /note= "Linker peptide."
;
; NAME/KEY: misc feature
; LOCATION: (1285-2322)
; OTHER INFORMATION: /note= "Estrogen binding domain."
;
US-08-564-264-2

Query Match 100.0%; Score 41; DB 3; Length 2322;
Best Local Similarity 100.0%; Pred. No. 1.1e-08;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGTTGATGCTGAGCCCCCATACTCTATTCCGAGTATGA 41
Db 1489 TTGTTGATGCTGAGCCCCCATACTCTATTCCGAGTATGA 1529

RESULT 5
US-08-076-726-16
; Sequence 16, Application US/08076726
; Patent No. 5464758
; GENERAL INFORMATION:
```

```
/ APPLICANT: Gossen, Manfred
/ APPLICANT: Bujard, Hermann
/ TITLE OF INVENTION: Tight Control of Gene Expression in
/ TITLE OF INVENTION: Eucaryotic Cells by Tetracycline-responsive Promoters
/ NUMBER OF SEQUENCES: 16
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Sterne, Kessler, Goldstein & Fox
/ STREET: 1100 New York Avenue, N.W.
/ CITY: Washington
/ STATE: District of Columbia
/ COUNTRY: United States of America
/ ZIP: 20005-3934
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/076,726
/ FILING DATE: 14-JUN-1993
/ CLASSIFICATION: 435
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Esmond, Robert W.
/ REGISTRATION NUMBER: 32,893
/ REFERENCE/DOCKET NUMBER: 0942.2490001
/ TELEPHONE: (202)371-2600
/ TELEFAX: (202)371-2540
/ TELEX: 248636 SSK
/ INFORMATION FOR SEQ ID NO: 16:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 4963 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: both
/ TOPOLOGY: both
/ US-08-076-726-16

Query Match 100.0%; Score 41; DB 1; Length 4963;
Best Local Similarity 100.0%; Pred. No. 1.4e-08;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGTTGGATGCTGAGCCCCCATACTCTATTCCGAGTATGA 41
Db 1420 TTGTTGGATGCTGAGCCCCCATACTCTATTCCGAGTATGA 1460

RESULT 6
US-08-260-452-9
/ Sequence 9, Application US/08260452
/ Patent No. 5650298
/ GENERAL INFORMATION:
/ APPLICANT: Gossen, Manfred
/ APPLICANT: Bujard, Hermann
/ APPLICANT: Salfeld, Jochen
/ APPLICANT: Voss, Jeffrey
/ TITLE OF INVENTION: Tight Control of Gene Expression in Eucaryotic
/ TITLE OF INVENTION: Cells by Tetracycline-responsive Promoters
/ NUMBER OF SEQUENCES: 10
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Lahive & Cockfield
/ STREET: 60 State Street
/ CITY: Boston
/ STATE: Massachusetts
/ COUNTRY: USA
/ ZIP: 02109-1875
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: ASCII text
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/260,452
/ FILING DATE:

US-09-933-267a-1_copy_306362_306402.rni
```

```
/ CLASSIFICATION: 435
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 08/076,327
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Giulio A. DeConti, Jr.
/ REGISTRATION NUMBER: 31,503
/ REFERENCE/DOCKET NUMBER: BBI-013CP
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (617) 227-7400
/ TELEFAX: (617) 227-5941
/ INFORMATION FOR SEQ ID NO: 9:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 4963 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: double
/ TOPOLOGY: circular
/ MOLECULE TYPE: DNA (genomic)
/ ORGANISM: Human cytomegalovirus
/ IMMEDIATE SOURCE:
/ CLONE: PUHD BGR4
/ US-08-260-452-9

Query Match 100.0%; Score 41; DB 1; Length 4963;
Best Local Similarity 100.0%; Pred. No. 1.4e-08;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGTTGGATGCTGAGCCCCCATACTCTATTCCGAGTATGA 41
Db 1420 TTGTTGGATGCTGAGCCCCCATACTCTATTCCGAGTATGA 1460

RESULT 7
US-08-481-970-9
/ Sequence 9, Application US/08481970
/ Patent No. 5859310
/ GENERAL INFORMATION:
/ APPLICANT: Gossen, Manfred
/ APPLICANT: Bujard, Hermann
/ APPLICANT: Salfeld, Jochen
/ APPLICANT: Voss, Jeffrey
/ TITLE OF INVENTION: Animal Transgenic for a Tetracycline-Controlled
/ NUMBER OF SEQUENCES: 10
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Lahive & Cockfield
/ STREET: 60 State Street, Suite 510
/ CITY: Boston
/ STATE: Massachusetts
/ COUNTRY: USA
/ ZIP: 02109-1875
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: ASCII text
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/481,970
/ FILING DATE:
/ CLASSIFICATION: 800
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 08/260,452
/ FILING DATE: 14-JUN-1994
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 08/076,327
/ FILING DATE: 14-JUN-1993
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Giulio A. DeConti, Jr.
/ REGISTRATION NUMBER: 31,503
/ REFERENCE/DOCKET NUMBER: BBI-013CP2
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (617) 227-7400
/ TELEFAX: (617) 227-5941
/ INFORMATION FOR SEQ ID NO: 9:

Transcript
```

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 4963 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Human cytomegalovirus
; IMMEDIATE SOURCE:
; CLONE: pUHD BGR4
;
US-08-481-970-9
;
Query Match 100.0%; Score 41; DB 2; Length 4963;
Best Local Similarity 100.0%; Pred. No. 1.4e-08;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGTTGGATGCTGAGCCCCCATACTCTATTCCGAGTATGA 41
Db 1420 TTGTTGGATGCTGAGCCCCCATACTCTATTCCGAGTATGA 1460

RESULT 8
US-08-897-719-9
; Sequence 9, Application US/08897719
; Patent No. 5922927
; GENERAL INFORMATION:
; APPLICANT: Gossen, Manfred
; APPLICANT: Bujard, Hermann
; APPLICANT: Salfeld, Jochen
; APPLICANT: Voss, Jeffrey
; TITLE OF INVENTION: Tight Control of Gene Expression in Eucaryotic
; TITLE OF INVENTION: Cells by Tetracycline-responsive Promoters
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lahive & Cockfield
; STREET: 60 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/897,719
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/260,452
; FILING DATE:
; FILING DATE:
; APPLICATION NUMBER: 08/076,327
; FILING DATE: 14-JUN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Giulio A. DeConti, Jr.
; REGISTRATION NUMBER: 31,503
; REFERENCE/DOCKET NUMBER: BBI-0113CP
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4963 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Human cytomegalovirus
; IMMEDIATE SOURCE:
; CLONE: pUHD BGR4
;
US-08-897-719-9
;
Query Match 100.0%; Score 41; DB 2; Length 4963;
Best Local Similarity 100.0%; Pred. No. 1.4e-08;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGTTGGATGCTGAGCCCCCATACTCTATTCCGAGTATGA 41
Db 1420 TTGTTGGATGCTGAGCCCCCATACTCTATTCCGAGTATGA 1460

RESULT 9
US-09-163-269-9
; Sequence 9, Application US/09163269
; Patent No. 6252136
; GENERAL INFORMATION:
; APPLICANT: Gossen, Manfred
; APPLICANT: Bujard, Hermann
; APPLICANT: Salfeld, Jochen
; APPLICANT: Voss, Jeffrey
; TITLE OF INVENTION: Animal Transgenic for a Tetracycline-Controlled
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lahive & Cockfield
; STREET: 60 State Street, Suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/163,269
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/481,970
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/076,327
; FILING DATE: 14-JUN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Giulio A. DeConti, Jr.
; REGISTRATION NUMBER: 31,503
; REFERENCE/DOCKET NUMBER: BBI-0113CP2
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4963 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Human cytomegalovirus
; IMMEDIATE SOURCE:
; CLONE: pUHD BGR4
;
US-09-163-269-9
;
Query Match 100.0%; Score 41; DB 3; Length 4963;
Best Local Similarity 100.0%; Pred. No. 1.4e-08;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGTTGGATGCTGAGCCCCCATACTCTATTCCGAGTATGA 41
Db 1420 TTGTTGGATGCTGAGCCCCCATACTCTATTCCGAGTATGA 1460

RESULT 10
```

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 4963 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Human cytomegalovirus
; IMMEDIATE SOURCE:
; CLONE: pUHD BGR4
;
US-08-481-970-9
;
Query Match 100.0%; Score 41; DB 2; Length 4963;
Best Local Similarity 100.0%; Pred. No. 1.4e-08;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGTTGGATGCTGAGCCCCCATACTCTATTCCGAGTATGA 41
Db 1420 TTGTTGGATGCTGAGCCCCCATACTCTATTCCGAGTATGA 1460

RESULT 8
US-08-897-719-9
; Sequence 9, Application US/08897719
; Patent No. 5922927
; GENERAL INFORMATION:
; APPLICANT: Gossen, Manfred
; APPLICANT: Bujard, Hermann
; APPLICANT: Salfeld, Jochen
; APPLICANT: Voss, Jeffrey
; TITLE OF INVENTION: Tight Control of Gene Expression in Eucaryotic
; TITLE OF INVENTION: Cells by Tetracycline-responsive Promoters
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lahive & Cockfield
; STREET: 60 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/897,719
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/260,452
; FILING DATE:
; FILING DATE:
; APPLICATION NUMBER: 08/076,327
; FILING DATE: 14-JUN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Giulio A. DeConti, Jr.
; REGISTRATION NUMBER: 31,503
; REFERENCE/DOCKET NUMBER: BBI-0113CP
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4963 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Human cytomegalovirus
; IMMEDIATE SOURCE:
; CLONE: pUHD BGR4
;
US-08-897-719-9
;
Query Match 100.0%; Score 41; DB 2; Length 4963;
Best Local Similarity 100.0%; Pred. No. 1.4e-08;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGTTGGATGCTGAGCCCCCATACTCTATTCCGAGTATGA 41
Db 1420 TTGTTGGATGCTGAGCCCCCATACTCTATTCCGAGTATGA 1460

RESULT 9
US-09-163-269-9
; Sequence 9, Application US/09163269
; Patent No. 6252136
; GENERAL INFORMATION:
; APPLICANT: Gossen, Manfred
; APPLICANT: Bujard, Hermann
; APPLICANT: Salfeld, Jochen
; APPLICANT: Voss, Jeffrey
; TITLE OF INVENTION: Animal Transgenic for a Tetracycline-Controlled
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lahive & Cockfield
; STREET: 60 State Street, Suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/163,269
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/481,970
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/076,327
; FILING DATE: 14-JUN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Giulio A. DeConti, Jr.
; REGISTRATION NUMBER: 31,503
; REFERENCE/DOCKET NUMBER: BBI-0113CP2
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4963 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Human cytomegalovirus
; IMMEDIATE SOURCE:
; CLONE: pUHD BGR4
;
US-09-163-269-9
;
Query Match 100.0%; Score 41; DB 3; Length 4963;
Best Local Similarity 100.0%; Pred. No. 1.4e-08;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGTTGGATGCTGAGCCCCCATACTCTATTCCGAGTATGA 41
Db 1420 TTGTTGGATGCTGAGCCCCCATACTCTATTCCGAGTATGA 1460

RESULT 10
```


US-09-281-674-9
; Sequence 9, Application US/09281674
; Patent No. 6783756
; GENERAL INFORMATION:
; APPLICANT: Gossen, Manfred
; Bujard, Hermann
; Salfeld, Jochen
; Voss, Jeffrey
; TITLE OF INVENTION: Methods for Regulating Gene Expression
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lahive & Cockfield
; STREET: 60 State Street, Suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII text
; CURRENT APPLICATION DATA:
; FILING DATE: 30-Mar-1999
; APPLICATION NUMBER: US/09/281,674
; FILING DATE: 30-Mar-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/479,306
; FILING DATE: <Unknown>
; APPLICATION NUMBER: 08/260,452
; FILING DATE: 14-JUN-1994
; APPLICATION NUMBER: 08/076,327
; FILING DATE: 14-JUN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Giulio A. DeConti, Jr.
; REGISTRATION NUMBER: 31,503
; REFERENCE/DOCKET NUMBER: BBI-013CP3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4963 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Human cytomegalovirus
; IMMEDIATE SOURCE:
; CLONE: pUHD BGR4
; SEQUENCE DESCRIPTION: SEQ ID NO: 9:
US-09-281-674-9
Query Match 100.0%; Score 41; DB 4; Length 4963;
Best Local Similarity 100.0%; Pred. No. 1.4e-08;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TTGTTGGATGCTGAGCCCCCATCTCTATTCCGAGTATGA 41
DB 1420 TTGTTGGATGCTGAGCCCCCATCTCTATTCCGAGTATGA 1460
RESULT 11
US-09-041-886-34
; Sequence 34, Application US/09041886
; Patent No. 6235872
; GENERAL INFORMATION:
; APPLICANT: Bredesen, Dale E.
; APPLICANT: Rabizadeh, Sharoz
; TITLE OF INVENTION: Proapoptotic Peptides, Dependence
; TITLE OF INVENTION: Polyptptides and Methods of Use
; NUMBER OF SEQUENCES: 72

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/041,886
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 2626
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 34:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6450 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 361..2146
US-09-041-886-34
Query Match 100.0%; Score 41; DB 3; Length 6450;
Best Local Similarity 100.0%; Pred. No. 1.5e-08;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TTGTTGGATGCTGAGCCCCCATCTCTATTCCGAGTATGA 41
DB 1315 TTGTTGGATGCTGAGCCCCCATCTCTATTCCGAGTATGA 1355
RESULT 12
US-08-453-998-1
; Sequence 1, Application US/08453998
; Patent No. 644438
; GENERAL INFORMATION:
; APPLICANT: CHAMRON, PIERRE
; APPLICANT: METZGER, DANIEL
; APPLICANT: WHITE, JOHN
; TITLE OF INVENTION: METHOD FOR THE PREPARATION OF A PROTEIN
; BY YEASTS USING AN INDUCIBLE SYSTEM, VECTORS AND
; CORRESPONDING TRANSFORMED STRAINS
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN, DARBY & CUSHMAN
; STREET: 1100 NEW YORK AVE., N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/453,998
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:

```
; APPLICATION NUMBER: US 07/161,064
; FILING DATE: 03-DEC-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: CHAPIN, MARLANA K.
; REGISTRATION NUMBER: 35,843
; REFERENCE/DOCKET NUMBER: 1037/98493
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-861-3711
; TELEFAX: 202-822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6450 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-453-998-1

Query Match 100.0%; Score 41; DB 4; Length 6450;
Best Local Similarity 100.0%; Pred. No. 1.5e-08;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TTGTTGGATGCTGAGCCCCCATACTCTATTCCGAGTATGA 41
Db 1315 TTGTTGGATGCTGAGCCCCCATACTCTATTCCGAGTATGA 1355

RESULT 13
US-10-027-983-3
; Sequence 3, Application US/10027983
; Patent No. 6617162
; GENERAL INFORMATION:
; APPLICANT: Kenneth W. Dobie
; APPLICANT: Mark P. Roach
; TITLE OF INVENTION: ANTISENSE MODULATION OF ESTROGEN RECEPTOR ALPHA EXPRESSION
; FILE REFERENCE: RTS-0340
; CURRENT APPLICATION NUMBER: US/10/027,983
; CURRENT FILING DATE: 2001-12-18
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 3
; LENGTH: 6450
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (361)...(2148)
; US-10-027-983-3

Query Match 100.0%; Score 41; DB 4; Length 6450;
Best Local Similarity 100.0%; Pred. No. 1.5e-08;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TTGTTGGATGCTGAGCCCCCATACTCTATTCCGAGTATGA 41
Db 1315 TTGTTGGATGCTGAGCCCCCATACTCTATTCCGAGTATGA 1355

RESULT 14
US-10-027-983-10
; Sequence 10, Application US/10027983
; Patent No. 6617162
; GENERAL INFORMATION:
; APPLICANT: Kenneth W. Dobie
; APPLICANT: Mark P. Roach
; TITLE OF INVENTION: ANTISENSE MODULATION OF ESTROGEN RECEPTOR ALPHA EXPRESSION
; FILE REFERENCE: RTS-0340
; CURRENT APPLICATION NUMBER: US/10/027,983
; CURRENT FILING DATE: 2001-12-18
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 10
; LENGTH: 8566
; TYPE: DNA
```

```
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: (1)...(2302)
; NAME/KEY: exon:exon junction
; LOCATION: (1491)...(1492)
; OTHER INFORMATION: exon 4:exon 5
; NAME/KEY: start codon
; LOCATION: (2302)...(2305)
; NAME/KEY: 3'UTR
; LOCATION: (2735)...(8566)
; NAME/KEY: exon:exon junction
; LOCATION: (3569)...(3570)
; OTHER INFORMATION: exon 8:exon 9
; NAME/KEY: exon:exon junction
; LOCATION: (3708)...(3709)
; OTHER INFORMATION: exon 9:exon 10
; NAME/KEY: exon:exon junction
; LOCATION: (3842)...(3843)
; OTHER INFORMATION: exon 10:exon 11
; NAME/KEY: exon:exon junction
; LOCATION: (4026)...(4027)
; OTHER INFORMATION: exon 11:exon 12
; US-10-027-983-10

Query Match 100.0%; Score 41; DB 4; Length 8566;
Best Local Similarity 100.0%; Pred. No. 1.6e-08;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TTGTTGGATGCTGAGCCCCCATACTCTATTCCGAGTATGA 41
Db 3431 TTGTTGGATGCTGAGCCCCCATACTCTATTCCGAGTATGA 3471

RESULT 15
US-09-270-767-8774
; Sequence 8774, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8774
; LENGTH: 334
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
; US-09-270-767-8774

Query Match 52.7%; Score 21.6; DB 4; Length 334;
Best Local Similarity 75.0%; Pred. No. 2.8;
Matches 27; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 2 TGTGTTGATGCTGAGCCCCCATACTCTATTCCGAGT 37
Db 31 TGGTGGACGGGAGCCCACTACCAAGTTCGAGT 66

Search completed: November 3, 2004, 09:46:52
Job time : 20.8325 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 3, 2004, 04:57:13 ; Search time 707.789 Seconds
(without alignments)
2110.838 Million cell updates/sec

Title: US-09-933-267A-1_COPY_64311_64351

Perfect score: 41

Sequence: 1 tctctcctgactttgtccgt.....aattgataagatctaatttg 41

Scoring table: IDENTITY NUC

Gapop 10_0 , Gapext 1.0

Searched: 32822875 seqs, 1821965908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:

1: gb_est1: *
2: gb_est2: *
3: gb_hic: *
4: gb_est3: *
5: gb_est4: *
6: gb_est5: *
7: gb_est6: *
8: gb_gss1: *
9: gb_gss2: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
C 1	25.2	61.5	651	6	CAB32730
C 2	24.8	60.5	711	1	AI725341
C 3	24	58.5	641	1	AV400698
C 4	23.4	57.1	540	5	BU779065
C 5	23.4	57.1	643	8	AZ381999
C 6	23	56.1	449	8	BH234952
C 7	23	56.1	798	6	CD765542
C 8	22.8	55.6	499	1	AL921335
C 9	22.8	55.6	633	8	CC063019
C 10	22.8	55.6	711	5	BP678983
C 11	22.8	55.6	1123	9	CL106907
C 12	22.8	55.6	1144	9	CL108627
C 13	22.6	55.1	600	8	AZ771688
C 14	22.6	55.1	630	8	BH181335
C 15	22.6	55.1	630	9	CNS07MZF
C 16	22.6	55.1	657	9	CE675920
C 17	22.6	55.1	835	8	CC321521
C 18	22.6	55.1	1226	9	CG770712
C 19	22.4	54.6	153	1	AA120330
C 20	22.4	54.6	179	1	AI790018
C 21	22.4	54.6	201	7	W12072
C 22	22.4	54.6	208	2	AW143047
C 23	22.4	54.6	221	1	AA467358
C 24	22.4	54.6	241	7	W12311

C 25	22.4	54.6	250	1	AA929314
C 26	22.4	54.6	250	1	AA177935
C 27	22.4	54.6	279	8	BH791401
C 28	22.4	54.6	280	5	EX638156
C 29	22.4	54.6	281	4	BM493236
C 30	22.4	54.6	286	4	BM403207
C 31	22.4	54.6	287	1	AI575934
C 32	22.4	54.6	288	1	AA198597
C 33	22.4	54.6	291	1	AV084080
C 34	22.4	54.6	294	2	BE134239
C 35	22.4	54.6	296	2	BE631235
C 36	22.4	54.6	300	7	CR470354
C 37	22.4	54.6	302	7	W11844
C 38	22.4	54.6	306	7	W35903
C 39	22.4	54.6	311	7	W62151
C 40	22.4	54.6	312	2	AW988519
C 41	22.4	54.6	314	7	W54541
C 42	22.4	54.6	319	1	AA060917
C 43	22.4	54.6	319	1	AA241978
C 44	22.4	54.6	319	7	W62065
C 45	22.4	54.6	320	2	BF524186

ALIGNMENTS

RESULT 1
CAB32730/c
LOCUS
DEFINITION
MCS009G07.150738 Ice plant Lambda Uni-Zap XR expression library, 5 days 0.5 M NaCl treatment, Crassulacean acid metabolism, phase I (2 AM). Mesembryanthemum crystallinum cDNA clone MCS009G07.5, mRNA sequence.
ACCESSION
CAB32730
VERSION
CAB32730.1
KEYWORDS
GI:26560495
SOURCE
Mesembryanthemum crystallinum (common iceplant)
ORGANISM
Mesembryanthemum crystallinum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Caryophyllales; Aizoaceae; Mesembryanthemum.
REFERENCE
1 (bases 1 to 651)
AUTHORS
Cushman, J.C.
TITLE
An expressed sequence tag database for the common ice plant, Mesembryanthemum crystallinum
JOURNAL
Unpublished (1997)
COMMENT
Contact: Cushman JC
Department of Biochemistry
University of Nevada
MS200, Reno, NV 89557-0014, USA
Tel: 775-784-1918
Fax: 775-784-1650
Email: jcushman@unr.edu
PCR Primers
FORWARD: T3 20mer
BACKWARD: T7 20mer
Plate: 009 row: G column: 07
Seq primer: T3 20mer
High quality sequence stop: 651.
Location/Qualifiers
1. 651
/organism="Mesembryanthemum crystallinum"
/mol_type="mRNA"
/db_xref="taxon:3544"
/clone="MCS009G07"
/tissue_type="leaf"
/dev_stage="five-week-old"
/clone_lib="Ice plant Lambda Uni-Zap XR expression library, 5 days 0.5 M NaCl treatment, Crassulacean acid metabolism, phase I (2 AM)."
/note="Vector: Lambda Uni-Zap XR, Bluescript SK-; Site 1: EcoRI; Site 2: XhoI; Library construction was performed according to Stratagene's recommended protocol for the

(20. 30 41)


```

ACCESSION BP678983
VERSION BP678983.1 GI:46026938
KEYWORDS EST.
SOURCE Xenopus laevis (African clawed frog)
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
Xenopodinae; Xenopus; Xenopus.
1 (bases 1 to 711)
Osada, S., Kitayama, A., Ueno, N. and Taira, M.
Expression analysis of genes which are expressed in the anterior
neuroectoderm of Xenopus embryos
Unpublished (2004)
CONTACT: Masanori Taira
Department of Biological Sciences
Graduate School of Science, University of Tokyo; CREST, Japan
Science and Technology Corporation, Japan
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan
Tel: 81-03-5841-4434
Fax: 81-03-5841-4434
Email: m.taira@biol.s.u-tokyo.ac.jp,
URL: http://www.shigen.nig.ac.jp/nbrp/xenopus/est/.
FEATURES             source
     location/Qualifiers
     1..711
     /organism="Xenopus laevis"
     /mol_type="mRNA"
     /db_xref="taxon:8355"
     /clone="XL430h22ex"
     /tissue_type="anterior neuroectoderm"
     /dev_stage="late gastrula (stage 12.5)"
     /clone_lib="Osada Taira anterior neuroectoderm (ANE)
     PCS105 cDNA library"

ORIGIN
Query Match          55.6%; Score 22.8; DB 5; Length 711;
Best Local Similarity 79.4%; Pred. No. 3.4e+02;
Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2 CTCCTCGACTTTGTCGTAATTCGTAAGATCT 35
    |||||
Db 581 CTCCTCACACATGTCGGAATTCGTAAGATCT 548

RESULT 11
LOCUS CL106907 1123 bp DNA linear GSS 05-JAN-2004
DEFINITION ISB1-4711.T7.1 ISB1 Xenopus tropicalis genomic clone ISB1-4711,
genomic survey sequence.
ACCESSION CL106907
VERSION CL106907.1 GI:40600542
KEYWORDS GSS.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
Xenopodinae; Xenopus; Silurana.
1 (bases 1 to 1123)
Kremitzki, C., Carter, J., McPherson, J., Warren, W., Graves, T.,
Mardis, E. and Wilson, R.
A physical map of the xenopus tropicalis genome
Unpublished (2003)
CONTACT: Richard K Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@watson.wustl.edu
Insert Length: 75000 Std Error: 0.00
Seq primer: T7 TAATACGACTCTATAGGG
Class: BAC ends
High quality sequence start: 14
High quality sequence stop: 759.
FEATURES             source
     location/Qualifiers
     1..1123
     /organism="Xenopus tropicalis"

ACCESSION BP678983
VERSION BP678983.1 GI:46026938
KEYWORDS EST.
SOURCE Xenopus laevis (African clawed frog)
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
Xenopodinae; Xenopus; Xenopus.
1 (bases 1 to 711)
Osada, S., Kitayama, A., Ueno, N. and Taira, M.
Expression analysis of genes which are expressed in the anterior
neuroectoderm of Xenopus embryos
Unpublished (2004)
CONTACT: Masanori Taira
Department of Biological Sciences
Graduate School of Science, University of Tokyo; CREST, Japan
Science and Technology Corporation, Japan
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan
Tel: 81-03-5841-4434
Fax: 81-03-5841-4434
Email: m.taira@biol.s.u-tokyo.ac.jp,
URL: http://www.shigen.nig.ac.jp/nbrp/xenopus/est/.
FEATURES             source
     location/Qualifiers
     1..711
     /organism="Xenopus laevis"
     /mol_type="mRNA"
     /db_xref="taxon:8355"
     /clone="XL430h22ex"
     /tissue_type="anterior neuroectoderm"
     /dev_stage="late gastrula (stage 12.5)"
     /clone_lib="Osada Taira anterior neuroectoderm (ANE)
     PCS105 cDNA library"

ORIGIN
Query Match          55.6%; Score 22.8; DB 5; Length 711;
Best Local Similarity 79.4%; Pred. No. 3.4e+02;
Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2 CTCCTCGACTTTGTCGTAATTCGTAAGATCT 35
    |||||
Db 581 CTCCTCACACATGTCGGAATTCGTAAGATCT 548

RESULT 11
LOCUS CL106907 1123 bp DNA linear GSS 05-JAN-2004
DEFINITION ISB1-4711.T7.1 ISB1 Xenopus tropicalis genomic clone ISB1-4711,
genomic survey sequence.
ACCESSION CL106907
VERSION CL106907.1 GI:40600542
KEYWORDS GSS.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
Xenopodinae; Xenopus; Silurana.
1 (bases 1 to 1123)
Kremitzki, C., Carter, J., McPherson, J., Warren, W., Graves, T.,
Mardis, E. and Wilson, R.
A physical map of the xenopus tropicalis genome
Unpublished (2003)
CONTACT: Richard K Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@watson.wustl.edu
Insert Length: 75000 Std Error: 0.00
Seq primer: T7 TAATACGACTCTATAGGG
Class: BAC ends
High quality sequence start: 14
High quality sequence stop: 759.
FEATURES             source
     location/Qualifiers
     1..1123
     /organism="Xenopus tropicalis"

```

```

/mol_type="genomic DNA"
/db_xref="taxon:8364"
/clone="ISB1-4711"
/clone_lib="ISB1"
/Note="Vector: pBelobAC11; ISB-1 Xenopus tropicalis BAC
Library Segment 1"

ORIGIN
Query Match          55.6%; Score 22.8; DB 9; Length 1123;
Best Local Similarity 79.4%; Pred. No. 3.6e+02;
Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 3 TCTCTGACTTTGTCGTAATTCGTAAGATCTA 36
    |||||
Db 471 TCTACTGACTTGCACCTCAATTATAGACCTA 504

RESULT 12
LOCUS CL108627/c 1144 bp DNA linear GSS 05-JAN-2004
DEFINITION ISB1-50C23.Sp6.1 ISB1 Xenopus tropicalis genomic clone ISB1-50C23,
genomic survey sequence.
ACCESSION CL108627
VERSION CL108627.1 GI:40602262
KEYWORDS GSS.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
Xenopodinae; Xenopus; Silurana.
1 (bases 1 to 1144)
Kremitzki, C., Carter, J., McPherson, J., Warren, W., Graves, T.,
Mardis, E. and Wilson, R.
A physical map of the xenopus tropicalis genome
Unpublished (2003)
CONTACT: Richard K Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@watson.wustl.edu
Insert Length: 75000 Std Error: 0.00
Seq primer: Sp6 ATTTAGGTGACACTATAG
Class: BAC ends
High quality sequence start: 107
High quality sequence stop: 633.
FEATURES             source
     location/Qualifiers
     1..1144
     /organism="Xenopus tropicalis"
     /mol_type="genomic DNA"
     /db_xref="taxon:8364"
     /clone="ISB1-50C23"
     /clone_lib="ISB1"
     /Note="Vector: pBelobAC11; ISB-1 Xenopus tropicalis BAC
     Library Segment 1"

ORIGIN
Query Match          55.6%; Score 22.8; DB 9; Length 1144;
Best Local Similarity 79.4%; Pred. No. 3.6e+02;
Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 3 TCTCTGACTTTGTCGTAATTCGTAAGATCTA 36
    |||||
Db 383 TCTACTGACTTGCACCTCAATTATAGACCTA 350

RESULT 13
LOCUS AZ771688 600 bp DNA linear GSS 16-FEB-2001
DEFINITION 1M0574H06F Mouse 10kb plasmid UUC1M library Mus musculus genomic
clone UUC1M0574H06 F, genomic survey sequence.
ACCESSION AZ771688
VERSION AZ771688.1 GI:12894212
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)

```



```

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source
1. .630
/organism="Schistosoma mansoni"
/mol_type="genomic DNA"
/strain="Puerto-Rican"
/db_xref="taxon:6183"
/clone="018BG06"
/clone_lib="SmBAC1"
/note="end : T7"
Location/Qualifiers

ORIGIN
Query Match 55.1%; Score 22.6; DB 9; Length 630;
Best Local Similarity 86.2%; Pred. No. 4e+02;
Matches 25; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 5 TCTGACTTGTCCGTAATAATGATAAGAT 33
292 TCAGAAATTTGTCGTAAATTTGATAAT 264

```

Search completed: November 3, 2004, 09:43:29
Job time : 711.123 secs

This Page Blank (uspto)

SECRET

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 3, 2004, 04:44:23 ; Search time 83.8636 Seconds
(without alignments)
2566.384 Million cell updates/sec

Title: US-09-933-267A-1_COPY_64311_64351

Perfect score: 41

Sequence: 1 tctctctgactttgtccgt.....aattgataagatctaatttg 41

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4134886 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_23Sep04:*

- 1: Geneseqn1980s:*
- 2: Geneseqn1990s:*
- 3: Geneseqn2000s:*
- 4: Geneseqn2001as:*
- 5: Geneseqn2001bs:*
- 6: Geneseqn2002as:*
- 7: Geneseqn2002bs:*
- 8: Geneseqn2003as:*
- 9: Geneseqn2003bs:*
- 10: Geneseqn2003cs:*
- 11: Geneseqn2003ds:*
- 12: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	41	100.0	110000	6	ABA90193_0 Human oes
2	41	100.0	110000	6	ABQ87681_0 Human oes
3	41	100.0	110000	8	ABX33717_0 Gene enco
4	39.4	96.1	110000	10	ADP81391_0 Partial g
5	23.8	58.0	1347	10	ADH82504_0 Enterococ
6	23.8	58.0	1368	4	AAS51295 Enterococ
7	23.8	58.0	1389	8	ACA18377 Prokaryot
8	23.8	58.0	1392	4	AAS52928 Enterococ
9	23.4	57.1	591	4	ABA63739 Human foe
10	23.4	57.1	591	4	AAI43851 Probe #12
11	23.4	57.1	591	4	AAK37973 Human bon
12	23.4	57.1	591	4	ABS37592 Human liv
13	23.4	57.1	591	6	ABS11972 Human gen
14	23	56.1	1360	12	AD116270 Human nuc
15	23	56.1	2069	10	ADB62733 Human cDN
16	22.4	54.6	658	3	AAA44181 Mouse sec
17	22.4	54.6	2709	8	ACA51279 Prokaryot
18	22.4	54.6	101786	3	AAP22293 BAC conta
19	22	53.7	32172	4	AAI03473 Human rep
20	21.8	53.2	1788	8	ABX34799 Human mdd
21	21.8	53.2	2274	5	AAS65768 DNA encod

C 22	21.8	53.2	2386	5	AAS86931	Human CDN
C 23	21.8	53.2	2621	10	ADB62795	Human CDN
C 24	21.8	53.2	2978	10	ADC37641	Human nuc
C 25	21.4	52.2	1541	8	ACA49041	Prokaryot
C 26	21.4	52.2	163350	6	AAAD46127	Human tum
C 27	21.2	51.7	16831	4	AAK73455	Human imm
C 28	21.2	51.7	89900	12	ADO79404	DPF3 regi
C 29	21.2	51.2	638	4	AAH08204	Human CDN
C 30	21.2	51.2	819	6	ABZ32168	Candida a
C 31	21.2	51.2	2356	4	AAH16613	Human CDN
C 32	21.2	51.2	3364	11	ADM01350	Human CDN
C 33	21.2	51.2	9540	12	ADO52562	Human imm
C 34	21.2	51.2	9540	12	ADP20074	Human imm
C 35	21.2	51.2	10872	6	ABK31389	Signal tr
C 36	21.2	51.2	10872	6	ABL70336	Chemical
C 37	21.2	51.2	10872	6	AAS61289	Human gen
C 38	20.8	50.7	375	5	ABV10419	Human pro
C 39	20.8	50.7	375	5	ABV01250	Human pro
C 40	20.8	50.7	409	5	ABV40557	Human pro
C 41	20.8	50.7	409	5	ABV31589	Human pro
C 42	20.8	50.7	423	8	ACF73919	Staphyloc
C 43	20.8	50.7	598	6	ABK53513	Human eos
C 44	20.8	50.7	951	10	ADF01416	Bacterial
C 45	20.8	50.7	1557	2	AAQ30973	HIV gp160

ALIGNMENTS

RESULT 1	ABA90193_0	LOCUS ABA90193 Accession ABA90193
WP	Sequence split into 5 fragments	Begin End
WP	Fragment Name	Begin End
WP	ABA90193_0	110000
WP	ABA90193_1	100001 210000
WP	ABA90193_2	200001 310000
WP	ABA90193_3	300001 410000
WP	ABA90193_4	400001 465237
ID	ABA90193 standard; DNA; 465237 BP.	
XX	ABA90193;	
XX		
DT	11-FEB-2002 (first entry)	
XX		
DE	Human oestrogen receptor alpha gene.	
XX		
KW	Human; oestrogen receptor alpha; ESR-alpha; ER; chromosome 6; Syne-2;	
KW	synaptic nuclei expressed gene 2; haplotype; cytostatic; osteopathic;	
KW	cardiant; vasotropic; gene therapy; vaccine; cancer; osteoporosis;	
KW	cardiovascular disease; oestrogen receptor; ds.	
OS	Homo sapiens.	
XX		
PN	WO200162969-A2.	
XX		
PD	30-AUG-2001.	
XX		
PF	20-FEB-2001; 2001WO-US005358.	
XX		
PR	22-FEB-2000; 2000US-0183756P.	
PR	20-OCT-2000; 2000US-00692414.	
PR	24-JAN-2001; 2001US-00768184.	
XX		
PA	(PEKE) PE CORP NY.	
XX		
PI	Kalush F, Cassel MJ, Hwang SS, Winn-Deen ES;	
DR	WPI: 2002-041152/05.	
DR	P-PSDB; AAG68251.	
XX		
PT	Novel variant of estrogen receptor alpha polypeptide useful for	
PT	determining the biological activity of a protein for high throughput	
PT	screening and for raising antibodies that elicit an immune response in	

A1 Frag

```

PT host.
XX Example; Fig 1 page 1-93; 333pp; English.
PS
CC The present invention describes an isolated peptide (I) consisting of an
CC amino acid sequence selected from: (a) the amino acid sequence of a
CC variant of the oestrogen receptor alpha (ESR-alpha) protein in AAG68251;
CC or (b) a fragment comprising at least 10 contiguous amino acids of the
CC protein in AAG68251. (I) has cytostatic, osteopathic, cardiant and
CC vasotropic activities, and can be used in gene therapy and vaccine
CC production. (II) is useful for identifying an agent that binds to (I), by
CC contacting (I) with an agent and assaying the contacted mixture to
CC determine whether a complex is formed with the agent bound to the
CC peptide. A polynucleotide (II), encoding (I), is useful in the
CC development of diagnostics and therapies for diseases and disorders
CC mediated/modulated by an oestrogen receptor (ER). (II) is also useful in
CC gene therapy for treating cancer, osteoporosis and cardiovascular
CC diseases. The human ESR-alpha gene is located on chromosome 6. The
CC present sequence represents the human ESR-alpha gene, which is given in
CC the exemplification of the present invention
XX
SQ Sequence 465237 BP; 133988 A; 89578 C; 93946 G; 147721 T; 0 U; 4 Other;
Query Match 100.0%; Score 41; DB 6; Length 110000;
Best Local Similarity 100.0%; Pred. No. 5.1e-06;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TCTCTCGACTTGTCCGTAATTAAGATCTAATTG 41
Db 64311 TCTCTCGACTTGTCCGTAATTAAGATCTAATTG 64351

RESULT 2
ABQ87681.0 LOCUS ABQ87681 Accession Abq87681
WP Sequence split into 5 fragments Begin End
WP Fragment Name 1 110000
WP ABQ87681_0 100001 210000
WP ABQ87681_1 200001 310000
WP ABQ87681_2 300001 410000
WP ABQ87681_3 400001 465237
WP ABQ87681_4
ID ABQ87681 standard; DNA; 465237 BP.
XX
AC ABQ87681;
XX
XX 18-SEP-2002 (first entry)
XX Human oestrogen receptor alpha gene.
XX
XX Human; oestrogen; receptor; oestrogen receptor alpha; cytostatic;
XX osteopathic; cardiant; cancer; osteoporosis; cardiovascular disorder;
XX chromosome 6q25.1; gene; ds.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FH variation replace(18783,T)
FT /*tag= a
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace(18937,C)
FT /*tag= a
FT /standard_name= "Single nucleotide polymorphism"
FT exon 18941..19032
FT /*tag= b
FT /number= 1
FT 19033..52817
FT /*tag= b
FT /number= 1
FT /cons_splice= (5'site:YES,3'site:NO)
FT variation replace(19034,C)
FT /*tag= a
FT /standard_name= "Single nucleotide polymorphism"
FT exon 52818..52940

```

```

FT variation
FT /*tag= b
FT /number= 2
FT replace(52877,A)
FT /*tag= a
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace(52901,A)
FT /*tag= a
FT /standard_name= "Single nucleotide polymorphism"
FT intron 52941..64149
FT /*tag= b
FT /number= 2
FT /cons_splice= (5'site:YES,3'site:NO)
FT exon 64150..64280
FT /*tag= b
FT /number= 3
FT 64281..166227
FT /*tag= b
FT /number= 3
FT /cons_splice= (5'site:YES,3'site:NO)
FT variation replace(64331,G)
FT /*tag= a
FT /standard_name= "Single nucleotide polymorphism"
FT exon 166228..166322
FT /*tag= b
FT /number= 4
FT 166323..168001
FT /*tag= b
FT /number= 4
FT replace(167950,G)
FT /*tag= a
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace(167989,G)
FT /*tag= a
FT /standard_name= "Single nucleotide polymorphism"
FT exon 168002..168120
FT /*tag= b
FT /number= 5
FT replace(168054,G)
FT /*tag= a
FT /standard_name= "Single nucleotide polymorphism"
FT intron 168121..169542
FT /*tag= b
FT /number= 5
FT /cons_splice= (5'site:YES,3'site:NO)
FT exon 169543..169825
FT /*tag= b
FT /number= 6
FT replace(169812,G)
FT /*tag= a
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace(169823,G)
FT /*tag= a
FT /standard_name= "Single nucleotide polymorphism"
FT intron 169826..169866
FT /*tag= b
FT /number= 6
FT /cons_splice= (5'site:YES,3'site:NO)
FT exon 169867..170678
FT /*tag= b
FT /number= 7
FT replace(170035,A)
FT /*tag= a
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace(170068,T)
FT /*tag= a
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace(170256,C)
FT /*tag= a
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace(170368,G)
FT /*tag= a
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace(170487,C)

```

```
FT /tag= a
FT /standard_name= "Single nucleotide polymorphism"
FT 170679. .204911
FT /tag= b
FT /number= 7
FT 204912. .205102
FT /tag= b
FT /number= 8
FT 205103. .242969
FT /tag= b
FT /number= 8
FT 242970. .243086
FT /tag= b
FT /number= 9
FT replace(243055,T)
FT /tag= a
FT /standard_name= "Single nucleotide polymorphism"
FT 243087. .306167
FT /tag= b
FT /number= 9
FT replace(243187,C)
FT /tag= a
FT /standard_name= "Single nucleotide polymorphism"
FT 306168. .306503
FT /tag= b
FT /number= 10
FT replace(306292,A)
FT /tag= a
FT /standard_name= "Single nucleotide polymorphism"
FT replace(306382,G)
FT /tag= a
FT /standard_name= "Single nucleotide polymorphism"
FT 306504. .373639
FT /tag= b
FT /number= 10
FT 373640. .373778
FT /tag= b
FT /number= 11
FT 373779. .422963
FT /tag= b
FT /number= 11
FT 422964. .423097
FT /tag= b
FT /number= 12
FT replace(423067,C)
FT /tag= a
FT /standard_name= "Single nucleotide polymorphism"
FT 423098. .456353
FT /tag= b
FT /number= 12
FT replace(423149,G)
FT /tag= a
FT /standard_name= "Single nucleotide polymorphism"
FT replace(423163,G)
FT /tag= a
FT /standard_name= "Single nucleotide polymorphism"
FT replace(423220,A)
FT /tag= a
FT /standard_name= "Single nucleotide polymorphism"
FT replace(423232,G)
FT /tag= a
FT /standard_name= "Single nucleotide polymorphism"
FT replace(423258,G)
FT /tag= a
FT /standard_name= "Single nucleotide polymorphism"
FT 456354. .456537
FT /tag= b
FT /number= 13
FT 456538. .460700
FT /tag= b
FT /number= 13
FT replace(459706,C)
FT /tag= a
```

```
FT variation
FT /standard_name= "Single nucleotide polymorphism"
FT replace(459832,A)
FT /tag= a
FT /standard_name= "Single nucleotide polymorphism"
FT replace(459913,G)
FT /tag= a
FT /standard_name= "Single nucleotide polymorphism"
FT replace(460024,G)
FT /tag= a
FT /standard_name= "Single nucleotide polymorphism"
FT replace(460056,T)
FT /tag= a
FT /standard_name= "Single nucleotide polymorphism"
FT replace(460159,G)
FT /tag= a
```

Query Match 100.0%; Score 41; DB 6; Length 110000;
Best Local Similarity 100.0%; Pred. No. 5.1e-06;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCTCTCTGACTTTGTCGTAATGATAAGATCTAATTG 41
Db 64311 TCTCTCTGACTTTGTCGTAATGATAAGATCTAATTG 64351

RESULT 3

ABX33717 0
WP Sequence split into 5 fragments LOCUS ABX33717 Accession Abx33717
WP Fragment Name Begin End
WP ABX33717_0 1 110000
WP ABX33717_1 100001 210000
WP ABX33717_2 200001 310000
WP ABX33717_3 300001 410000
WP ABX33717_4 400001 465237
ID ABX33717 standard; DNA; 465237 BP.

XX AC ABX33717;

XX XX (26-FEB-2003 (first entry))

XX Gene encoding human oestrogen receptor alpha protein (ESR1).

Human; oestrogen receptor alpha; ESR1; cancer; osteoporosis;
cardiovascular disorder; variant oestrogen receptor; ESR1 haplotype;
ESR1 polymorphism detection; cytostatic; osteopathic; cardiant;
Chromosome 6q25.1; gene; single nucleotide polymorphism; SNP; ds.

XX OS Homo sapiens.

XX Key Location/Qualifiers

FH variation replace(18783,T)

FT /tag= ao

FT /standard_name= "SNP"

FT variation replace(18937,C)

FT /tag= ap

FT /standard_name= "SNP"

FT exon 18941. .19032

FT /tag= a

FT /number= 1G

FT 19033. .52817

FT /tag= b

FT /number= 1G

FT variation replace(19034,C)

FT /tag= aq

FT /standard_name= "SNP"

FT exon 52818. .52940

FT /tag= C

FT /number= 1F

FT variation replace(52877,A)

FT /tag= an

FT /standard_name= "SNP"

FT variation replace(52901,A)

FT /tag= am

FT /*tag= bf
 FT /standard_name= "SNP"
 FT replace(460159,G)
 FT /*tag= bg
 FT /standard_name= "SNP"
 FT replace(460553,C)
 FT /*tag= bh
 FT /standard_name= "SNP"

Query Match 100.0%; Score 41; DB 8; Length 110000;
 Best Local Similarity 100.0%; Pred. No. 5.1e-06;
 Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCTCTCTGACTTTGTCGGTAAATTGATAAGATCTAATTG 41
 |||||
 Db 64311 TCTCTCTGACTTTGTCGGTAAATTGATAAGATCTAATTG 64351

RESULT 4

ADB81391_0
 WP Sequence split into 4 fragments LOCUS ADB81391 Accession Adb81391
 WP Fragment Name Begin End
 WP ADB81391_0 1 110000
 WP ADB81391_1 100001 210000
 WP ADB81391_2 200001 310000
 WP ADB81391_3 300001 392000
 ID ADB81391 standard; DNA; 392000 BP.

XX AC ADB81391;

XX XX 04-DEC-2003 (first entry)

XX DE Partial genomic DNA sequence of the human oestrogen receptor alpha mRNA.
 XX gene; human; ds; oestrogen receptor alpha; ESR-alpha;
 XX oestrogen receptor 1; ESR1; NR3A1; bone maintenance;
 XX cardiovascular system; cancer; gene therapy; hyperproliferative disease;
 XX inflammation; tumour formation; infection; cytostatic; antiinflammatory;
 XX antimicrobial.

XX OS Homo sapiens.

XX XX W02003052072-A2.

XX PD 26-JUN-2003.

XX PF 13-DEC-2002; 2002WO-US040083.

XX PR 18-DEC-2001; 2001US-00027983

XX PA (ISIS-) ISIS PHARM INC.

XX PI Dobie KW, Roach MP;

XX PS WPI; 2003-577322/54.

XX New antisense compound targeted to nucleic acid encoding estrogen
 PT receptor alpha and inhibiting expression of estrogen receptor alpha,
 PT useful for treating a disease or condition e.g. a hyperproliferative
 PT disease.

XX Example 15; Page 99-210; 232pp; English.

XX This invention relates to human oestrogen receptor alpha (ESR-alpha), and
 CC the novel antisense oligonucleotides that modulate its expression. The
 CC oestrogen receptor alpha protein is also known as oestrogen receptor 1,
 CC ESR1, and NR3A1. Oestrogen, the steroid hormone ligand of ESR-alpha, is
 CC important for bone maintenance and plays a protective role in the
 CC cardiovascular system, as well as being required for normal sexual
 CC maturation through promoting growth and differentiation. Splice variants
 CC of ESR-alpha, however, have been associated with various cancers
 CC including the breast and pituitary. Accordingly, antisense
 CC oligonucleotides that inhibit the expression of ESR-alpha in cells or

CC tissues can be used in gene therapy to treat conditions such as
 CC hyperproliferative disease, inflammation, tumour formation and to prevent
 CC or delay infection. As such, the present invention describes these
 CC antisense oligos as having cytostatic, antiinflammatory and antimicrobial
 CC activities. This polynucleotide is the partial genomic sequence of the
 CC human oestrogen receptor alpha DNA of the invention.

XX SQ Sequence 392000 BP; 113320A; 73484C; 75996G; 117294T; 0U; 119060Other;

Query Match 96.1%; Score 39.4; DB 10; Length 110000;
 Best Local Similarity 97.6%; Pred. No. 2.2e-05;
 Matches 40; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCTCTCTGACTTTGTCGGTAAATTGATAAGATCTAATTG 41
 |||||
 Db 12773 TCTCTCTGACTTTGTCGGTAAATTGATAAGATCTAATTG 12813

RESULT 5

ADH82504/C
 ID ADH82504 standard; DNA; 1347 BP.

XX AC ADH82504;

XX DT 22-APR-2004 (first entry)

XX DE Enterococcus faecalis polynucleotide #389.

XX KW Enterococcus faecalis infection; transcription regulatory element;
 XX antibacterial; gene; ds.

XX OS Enterococcus faecalis.

XX PN US6617156-B1.

XX PD 09-SEP-2003.

XX PF 13-AUG-1998; 98US-00134000.

XX PR 15-AUG-1997; 97US-0055778P.

XX PA (DOUC/) DOUCETTE-STAMM L A.

XX PI (BUSH/) BUSH D.

XX Doucette-Stamm LA, Bush D;

XX DR WPI; 2003-895394/82.

XX DR P-PSDB; ADH85909.

XX PT New nucleic acid comprising a sequence encoding an Enterococcus faecalis
 PT polypeptide, useful for preparing a composition for diagnosing or
 PT treating E. faecalis infection.

XX PS Disclosure; SEQ ID NO 389; 193pp; English.

XX The invention relates to Enterococcus faecalis polynucleotides and
 CC polypeptides. The invention also relates to a recombinant expression
 CC vector comprising a polynucleotide operably linked to a transcription
 CC regulatory element, a cell comprising a recombinant vector, a method for
 CC producing an E. faecalis polypeptide, an isolated nucleic acid comprising
 CC a sequence not given in the specification, a recombinant vector
 CC comprising the nucleic acid and a cell comprising the recombinant vector.
 CC The polynucleotides can be used to detect the presence of E. faecalis in
 CC a sample. The sequences are useful for preparing a composition for
 CC diagnosing or treating Enterococcus faecalis infection. This sequence
 CC represents an E. faecalis polynucleotide of the invention.

XX SQ Sequence 1347 BP; 322 A; 217 C; 328 G; 480 T; 0 U; 0 Other;

Query Match 58.0%; Score 23.8; DB 10; Length 1347;
 Best Local Similarity 80.0%; Pred. No. 16;
 Matches 28; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 3 TCTCCTGACCTTGTCCGTAATTTGATAAGATCTAA 37
||||| | | | | | | | | | | | | | | | | | | | |
Db 1151 TCTCCTTTTCATGTGCGTAATTTGATAATTCCTAA 1117

RESULT 6
AAS51295/c
ID AAS51295 standard; DNA; 1368 BP.
XX AC AAS51295;
XX DT 13-FEB-2002 (first entry)
XX DE Enterococcus faecalis DNA for cellular proliferation protein #72.
XX KW Antisense; ds; prokaryotic cellular proliferation gene; antibiotic;
XX KW antibacterial; drug design.
XX OS Enterococcus faecalis.
XX PN WO200170955-A2.
XX PD 27-SEP-2001.
XX PF 21-MAR-2001; 2001WO-US009180.
XX PR 21-MAR-2000; 2000US-0191078P.
XX PR 23-MAY-2000; 2000US-0206848P.
XX PR 26-MAY-2000; 2000US-0207727P.
XX PR 23-OCT-2000; 2000US-0242578P.
XX PR 27-NOV-2000; 2000US-0253625P.
XX PR 22-DEC-2000; 2000US-0257931P.
XX PR 16-FEB-2001; 2001US-0269308P.
XX PA (ELIT-) ELITRA PHARM INC.
XX PI Haselbeck R, Ohlssen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
XX PI Yamamoto RT, Xu HH;
XX WPI; 2001-611495/70.
XX P-PSDB; AAU33436.
XX New polynucleotides for the identification and development of
XX antibiotics, comprise sequences of antisense nucleic acids.
XX Claim 27; SEQ ID NO 3877; 51lpp; English.
XX The invention relates to antisense inhibitors of genes essential to
XX prokaryotic cellular proliferation, their use in identifying the genes,
XX their use in the discovery of novel antibiotics, the essential genes
XX themselves and the encoded proteins. The prokaryotes used are Escherichia
XX coli, Staphylococcus aureus, Salmonella typhi, Klebsiella pneumoniae,
XX Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also
XX useful for the identification of potential new targets for antibiotic
XX development. The antisense nucleic acids can also be used to identify
XX proteins used in proliferation, to express these proteins, and to obtain
XX antibodies capable of binding to the expressed proteins. The proteins can
XX be used to screen compounds in rational drug discovery programmes. The
XX antisense nucleic acid sequence is also useful to screen for homologous
XX nucleic acids which are required for cell proliferation in a wide variety
XX of organisms. The present sequence encodes an essential prokaryotic
XX cellular proliferation protein. Note: The sequence data for this patent
XX did not form part of the printed specification, but was obtained in
XX electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 1368 BP; 330 A; 220 C; 332 G; 486 T; 0 U; 0 Other;

Query Match 58.0%; Score 23.8; DB 4; Length 1368;
Best Local Similarity 80.0%; Pred. No. 16;
Matches 28; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 3 TCTCCTGACCTTGTCCGTAATTTGATAAGATCTAA 37
||||| | | | | | | | | | | | | | | | | | | | |
Db 1151 TCTCCTTTTCATGTGCGTAATTTGATAATTCCTAA 1117

RESULT 7
ACA18377/c
ID ACA18377 standard; DNA; 1389 BP.
XX AC ACA18377;
XX DT 19-JUN-2003 (first entry)
XX DE Prokaryotic essential gene #34.
XX KW Antisense; ds; prokaryotic essential gene; cell proliferation;
XX KW drug design; gene.
XX OS Enterococcus faecalis.
XX PN WO200277183-A2.
XX PD 03-OCT-2002.
XX PF 21-MAR-2002; 2002WO-US009107.
XX PR 21-MAR-2001; 2001US-00815242.
XX PR 06-SEP-2001; 2001US-00948993.
XX PR 25-OCT-2001; 2001US-0342923P.
XX PR 08-FEB-2002; 2002US-00072851.
XX PR 06-MAR-2002; 2002US-0362699P.
XX PA (ELIT-) ELITRA PHARM INC.
XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlssen KL, Zyskind JW;
XX PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX WPI; 2003-029926/02.
XX P-PSDB; ABU14507.
XX New antisense nucleic acids, useful for identifying proteins or screening
XX for homologous nucleic acids required for cellular proliferation to
XX isolate candidate molecules for rational drug discovery programs.
XX Claim 14; SEQ ID NO 6247; 1766pp; English.
XX The invention relates to an isolated nucleic acid comprising any one of
XX the 6213 antisense sequences given in the specification where expression
XX of the nucleic acid inhibits proliferation of a cell. Also included are:
XX (1) a vector comprising a promoter operably linked to the nucleic acid
XX encoding a polypeptide whose expression is inhibited by the antisense
XX nucleic acid; (2) a host cell containing the vector; (3) an isolated
XX polypeptide or its fragment whose expression is inhibited by the
XX antisense nucleic acid; (4) an antibody capable of specifically binding
XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
XX proliferation or the activity of a gene in an operon required for
XX proliferation; (7) identifying a compound that influences the activity of
XX the gene product or that has an activity against a biological pathway
XX required for proliferation, or that inhibits cellular proliferation; (8)
XX identifying a gene required for cellular proliferation or the biological
XX pathway in which a proliferation-required gene or its gene product lies
XX or a gene on which the test compound that inhibits proliferation of an
XX organism acts; (9) manufacturing an antibiotic; (10) profiling a
XX compound's activity; (11) a culture comprising strains in which the gene
XX product is overexpressed or underexpressed; (12) determining the extent
XX to which each of the strains is present in a culture or collection of
XX strains; or (13) identifying the target of a compound that inhibits the
XX proliferation of an organism. The antisense nucleic acids are useful for
XX identifying proteins or screening for homologous nucleic acids required
XX for cellular proliferation to isolate candidate molecules for rational
XX drug discovery programs, or for screening homologous nucleic acids
XX required for proliferation in cells other than S. aureus, S. typhimurium,
XX K. pneumoniae or P. aeruginosa. The present sequence is one of the target
XX prokaryotic essential genes. Note: The sequence data for this patent did

CC not form part of the printed specification; but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1389 BP; 337 A; 224 C; 337 G; 491 T; 0 U; 0 Other;
Query Match 58.0%; Score 23.8; DB 8; Length 1389;
Best Local Similarity 80.0%; Pred. No. 17;
Matches 28; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
QY 3 TCTCCTGACCTTGTCCGTAATTCATAGATCTTAA 37
||||| ||||| ||||| ||||| |||||
Db 1196 TCTCCTTTCATGTCCGTAATTCATAGATCTTAA 1162

RESULT 8
AAS52928/C
ID AAS52928 standard; DNA; 1392 BP.
XX AC AAS52928;
XX DT 13-FEB-2002 (first entry)
XX DE Enterococcus faecalis DNA for cellular proliferation protein #356.
XX KW Antisense; ds; prokaryotic cellular proliferation gene; antibiotic;
XX KW antibacterial; drug design.
XX OS Enterococcus faecalis.
XX PN WO200170955-A2.
XX PD 27-SEP-2001.
XX PF 21-MAR-2001; 2001WO-US009180.
XX PR 21-MAR-2000; 2000US-0191078P.
XX PR 23-MAY-2000; 2000US-0206848P.
XX PR 26-MAY-2000; 2000US-0207727P.
XX PR 23-OCT-2000; 2000US-0242578P.
XX PR 27-NOV-2000; 2000US-0253625P.
XX PR 22-DEC-2000; 2000US-0257931P.
XX PR 16-FEB-2001; 2001US-0269308P.
XX PA (ELIT-) ELITRA PHARM INC.
XX PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
XX PI Yamamoto RT, Xu HH;
XX P-PSDB; AAU35069.
XX PT New polynucleotides for the identification and development of
XX PT antibiotics, comprise sequences of antisense nucleic acids.
XX PS Claim 27; SEQ ID NO 6565; 511pp; English.
XX CC The invention relates to antisense inhibitors of genes essential to
XX CC prokaryotic cellular proliferation, their use in identifying the genes,
XX CC their use in the discovery of novel antibiotics, the essential genes
XX CC themselves and the encoded proteins. The prokaryotes used are Escherichia
XX CC coli, Staphylococcus aureus, Salmonella typhi, Klebsiella pneumoniae,
XX CC Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also
XX CC useful for the identification of potential new targets for antibiotic
XX CC development. The antisense nucleic acids can also be used to identify
XX CC proteins used in proliferation, to express these proteins, and to obtain
XX CC antibodies capable of binding to the expressed proteins. The proteins can
XX CC be used to screen compounds in rational drug discovery programmes. The
XX CC antisense nucleic acid sequence is also useful to screen for homologous
XX CC nucleic acids which are required for cell proliferation in a wide variety
XX CC of organisms. The present sequence encodes an essential prokaryotic
XX CC cellular proliferation protein. Note: The sequence data for this patent
XX CC did not form part of the printed specification, but was obtained in

CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1392 BP; 341 A; 223 C; 337 G; 491 T; 0 U; 0 Other;
Query Match 58.0%; Score 23.8; DB 4; Length 1392;
Best Local Similarity 80.0%; Pred. No. 17;
Matches 28; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
QY 3 TCTCCTGACCTTGTCCGTAATTCATAGATCTTAA 37
||||| ||||| ||||| ||||| |||||
Db 1196 TCTCCTTTCATGTCCGTAATTCATAGATCTTAA 1162

RESULT 9
ABA63739
ID ABA63739 standard; DNA; 591 BP.
XX AC ABA63739;
XX DT 01-FEB-2002 (first entry)
XX DE Human foetal liver single exon nucleic acid probe #12044.
XX KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX KW Homo sapiens.
XX OS Homo sapiens.
XX PN WO200157277-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US000669.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234887P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-483447/52.
XX PT Human genome-derived single exon nucleic acid probes useful for analyzing
XX PT gene expression in human fetal liver.
XX PS Claim 1; SEQ ID NO 12044; 639pp + Sequence Listing; English.
XX CC The invention relates to a single exon nucleic acid probe for measuring
XX CC human gene expression in a sample derived from human foetal liver. The
XX CC single exon nucleic acid probes may be used for predicting, measuring and
XX CC displaying gene expression in samples derived from human fetal liver. The
XX CC present sequence is a single exon nucleic acid probe of the invention.
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 591 BP; 188 A; 89 C; 98 G; 216 T; 0 U; 0 Other;
Query Match 57.1%; Score 23.4; DB 4; Length 591;
Best Local Similarity 73.2%; Pred. No. 21;
Matches 30; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
QY 1 TCTCCTGACCTTGTCCGTAATTCATAGATCTTAA 41
||||| ||||| ||||| ||||| |||||
Db 80 TCTCTACTGACTTCTCTTTAATTCATAGTCTTAA 120

```
RESULT 10
AAI43851
ID AAI43851 standard; DNA; 591 BP.
XX
XX AC AAI43851;
XX
XX DT 17-OCT-2001 (first entry)
XX
XX DE Probe #12537 used to measure gene expression in human placenta sample.
XX
XX KW Probe; microarray; human; placenta; antenatal diagnosis;
XX KW genetic disorder; ss.
XX
XX OS Homo sapiens.
XX
XX PN WO200157272-A2.
XX
XX PD 09-AUG-2001.
XX
XX PF 30-JAN-2001; 2001WO-US000663.
XX
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488997/53.
XX
XX PT Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human placenta.
XX
XX PS Claim 25; SEQ ID NO 12537; 654pp; English.
XX
XX CC The present invention relates to single exon nucleic acid probes (SENP).
XX CC The present sequence is one such probe. The probes are useful for
XX CC producing a microarray for predicting, measuring and displaying gene
XX CC expression in samples derived from human placenta. The probes are useful
XX CC for antenatal diagnosis of human genetic disorders
XX
XX SQ Sequence 591 BP; 188 A; 89 C; 98 G; 216 T; 0 U; 0 Other;

Query Match 57.1%; Score 23.4; DB 4; Length 591;
Best Local Similarity 73.2%; Pred. No. 21;
Matches 30; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 1 TCTCTCCTGACTTTGCGTAAATTGATAGATCTTAATTG 41
||| ||||| ||| ||| ||| ||| ||| ||| ||| |||
DB 80 TCTCTACTGACTTCTCTTAATTAGCTTAGTGTAAATTG 120

RESULT 11
AAK37973
ID AAK37973 standard; DNA; 591 BP.
XX
XX AC AAK37973;
XX
XX DT 06-NOV-2001 (first entry)
XX
XX DE Human bone marrow expressed single exon probe SEQ ID NO: 12530.
XX
XX KW Human; bone marrow expressed exon; gene expression analysis; probe;
XX KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX
XX OS Homo sapiens.
XX
XX PN WO200157276-A2.

Query Match 57.1%; Score 23.4; DB 4; Length 591;
Best Local Similarity 73.2%; Pred. No. 21;
Matches 30; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 1 TCTCTCCTGACTTTGCGTAAATTGATAGATCTTAATTG 41
||| ||||| ||| ||| ||| ||| ||| ||| ||| |||
DB 80 TCTCTACTGACTTCTCTTAATTAGCTTAGTGTAAATTG 120

RESULT 12
ABS37592
ID ABS37592 standard; DNA; 591 BP.
XX
XX AC ABS37592;
XX
XX DT 25-FEB-2003 (first entry)
XX
XX DE Human liver single exon probe, SEQ ID No 12582.
XX
XX KW Human; single exon nucleic acid probe; liver; cirrhosis;
XX KW hyperlipoproteinaemia; hyperlipidaemia; hypercholesterolaemia;
XX KW coronary heart disease; ss.
XX
XX OS Homo sapiens.
XX
XX PN WO200157273-A2.
XX
XX PD 09-AUG-2001.
XX
XX PF 30-JAN-2001; 2001WO-US000664.
XX
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX
XX PA (MOLE-) MOLECULAR DYNAMICS INC.

Query Match 57.1%; Score 23.4; DB 4; Length 591;
Best Local Similarity 73.2%; Pred. No. 21;
Matches 30; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 1 TCTCTCCTGACTTTGCGTAAATTGATAGATCTTAATTG 41
||| ||||| ||| ||| ||| ||| ||| ||| ||| |||
DB 80 TCTCTACTGACTTCTCTTAATTAGCTTAGTGTAAATTG 120
```

XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-48898/53.
XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human adult liver.
XX
XX Claim 1; SEQ ID NO 12582; 658pp; English.
XX
XX The invention relates to a single exon nucleic acid probe (SENP) (I) for
XX measuring human gene expression in a sample derived from human adult
XX liver, comprising one of 13109 defined nucleotide sequences given in the
XX specification (or complements/ fragments). The probe hybridizes at high
XX stringency to a nucleic acid molecule expressed in the human adult liver.
XX (I) may be used for predicting, measuring and displaying gene expression
XX in samples derived from human adult liver. The genes identified may be
XX involved in genetic liver diseases such as cirrhosis,
XX hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is
XX associated with coronary heart disease. ABS25011-ABS51005 represent human
XX liver single exon nucleic acid probes of the invention. Note: The
XX sequence information for this patent does not appear in the printed
XX specification but was obtained in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 591 BP; 188 A; 89 C; 98 G; 216 T; 0 U; 0 Other;
XX
XX Query Match 57.1%; Score 23.4; DB 4; Length 591;
XX Best Local Similarity 73.2%; Pred. No. 21;
XX Matches 30; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
XX
XX QY 1 TCTCTCTGACTTTGTCGTAATTCGATAAGATCTAATTG 41
XX 80 TCTCTAGACTTCTCTTTAATAGCTTAGCTGTTAATTG 120
XX
XX RESULT 13
XX ABS11972
XX ID ABS11972 standard; DNA; 591 BP.
XX
XX AC ABS11972;
XX
XX DT 19-AUG-2002 (first entry)
XX
XX DE Human genome-derived single exon probe from lung SEQ ID NO 11963.
XX
XX Human; ds; single exon probe; asthma; lung cancer; COPD; ILD;
XX chronic obstructive pulmonary disease; interstitial lung disease;
XX familial idiopathic pulmonary fibrosis; neurofibromatosis;
XX tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
XX Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
XX pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;
XX pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
XX primary ciliary dyskinesia; pulmonary hypertension;
XX hyaline membrane disease.
XX
XX OS Homo sapiens.
XX
XX XX WO200186003-A2.
XX
XX PD 15-NOV-2001.
XX
XX PF 30-JAN-2001; 2001WO-US000665.
XX
XX PR 04-FEB-2000; 2000US-0180312P.
XX 26-MAY-2000; 2000US-0207456P.
XX 30-JUN-2000; 2000US-00608408.
XX 03-AUG-2000; 2000US-00632366.
XX 21-SEP-2000; 2000US-0234687P.
XX 27-SEP-2000; 2000US-0236359P.
XX 04-OCT-2000; 2000GB-00024263.
XX
XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2002-114183/15.
XX Spatially-addressable set of single exon nucleic acid probes, used to
XX measure gene expression in human lung samples.
XX
XX Claim 1; SEQ ID NO 11963; 634pp; English.
XX
XX The invention relates to a spatially-addressable set of single exon
XX nucleic acid probes for measuring gene expression in a sample derived
XX from human lung comprising single exon nucleic acid probes having one of
XX 12614 nucleic acid sequences mentioned in the specification, or their
XX complements or the 13387 open reading frames derived from the 12614
XX probes. Also included are a microarray comprising the novel set of probes
XX; the novel set of probes which hybridise at high stringency to a nucleic
XX acid expressed in the human lung; measuring gene expression in a sample
XX derived from human lung, comprising (a) contacting the array with a
XX collection of detectably labeled nucleic acids derived from human lung
XX mRNA, and (b) measuring the label detectably bound to each probe of the
XX array; identifying exons in a eukaryotic genome, comprising (a)
XX algorithmically predicting at least one exon from genomic sequences of
XX the eukaryote; and (b) detecting specific hybridisation of detectably
XX labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
XX having a fragment identical to the predicted exon, the probe is included
XX in the above mentioned microarray; assigning exons to a single gene,
XX comprising (a) identifying exons from genomic sequence by the method
XX above and (b) measuring the expression of each of the exons in several
XX tissues and/or cell types using hybridisation to a single exon
XX microarrays having a probe with the exon, where a common pattern of
XX expression of the exons in the tissues and/or cell types indicates that
XX the exons should be assigned to a single gene; a peptide comprising one
XX of 12011 sequences, mentioned in the specification, or encoded by the
XX probes/open reading frames (ORF). The probes are used for gene expression
XX analysis, and for identifying exons in a gene, particularly using human
XX lung derived mRNA and for the study of lung diseases such as asthma, lung
XX cancer, chronic obstructive pulmonary disease (COPD), interstitial lung
XX disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis,
XX tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-
XX Pudlak syndrome, sarcoidosis, pulmonary haemosiderosis, pulmonary
XX histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis,
XX Karagener syndrome, fibrocystic pulmonary dysplasia, primary ciliary
XX dyskinesia, pulmonary hypertension and hyaline membrane disease. The
XX present sequence is a single exon probe of the invention. Note: The
XX sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 591 BP; 188 A; 89 C; 98 G; 216 T; 0 U; 0 Other;
XX
XX Query Match 57.1%; Score 23.4; DB 6; Length 591;
XX Best Local Similarity 73.2%; Pred. No. 21;
XX Matches 30; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
XX
XX QY 1 TCTCTCTGACTTTGTCGTAATTCGATAAGATCTAATTG 41
XX 80 TCTCTAGACTTCTCTTTAATAGCTTAGCTGTTAATTG 120
XX
XX Db
XX
XX RESULT 14
XX ADI16270/c
XX ID ADI16270 standard; DNA; 1360 BP.
XX
XX AC ADI16270;
XX
XX DT 22-APR-2004 (first entry)
XX
XX DE Human nucleic acid-associated protein (NAAP) coding sequence #5.
XX human; nucleic acid-associated protein; NAAP; autoimmune disorder;
XX inflammatory disorder; AIDS; allergy; infection; metabolic disorder;
XX obesity; reproductive disorder; infertility; neurological disorder;
XX

KW Parkinson's disease; Alzheimer's disease; cardiovascular disorder;
KW myocardial infarction; hypertension; eye disorder;
KW cell proliferative disease; cancer; ds; gene.
XX
OS Homo sapiens.
XX WO2003094848-A2.
XX PN
XX PD 20-NOV-2003.
XX
PF 09-MAY-2003; 2003WO-US014450.
XX
XX 10-MAY-2002; 2002US-0379843P.
PR 24-MAY-2002; 2002US-0383457P.
PR 31-MAY-2002; 2002US-0384699P.
PR 06-JUN-2002; 2002US-0387265P.
XX
XX (INCY-) INCYTE CORP.
PA
XX Kable AE, Elliott VS, Tran UK, Ramkumar J, Marquis JP, Chawla NK,
PI Richardson TW, Bulloch SA, Khare R, Lee SY, Lal PG, Tang YT, Yue H;
PI Swarnakar A, Becha SD, Hafalia AJA, Chang H, Baughn MR, Borowsky ML;
PI Gietzen Ku, He A, Forsythe JG, Sprague MW, Blake JJ, Warren BA;
PI Mason PM, Ison CH, Lindquist EA, Wilson AD, Jin P;
XX WPI; 2004-011999/01.
DR P-PSDB; ADI16220.
XX
XX New human nucleic acid associated proteins and polynucleotides, useful
PT for diagnosing, preventing or treating diseases or conditions associated
PT with aberrant protein expression, e.g. cancer, AIDS, atherosclerosis or
PT stroke.
XX
PS Claim 5; SEQ ID NO 55; 400pp; English.
XX
XX The invention comprises the amino acid and coding sequences of human
CC nucleic acid-associated proteins (NAAP). The DNA and protein sequences of
CC the invention are useful in diagnosing, preventing and treating
CC diseases/conditions associated with altered expression of NAAP, such as:
CC autoimmune/inflammatory disorders (e.g. AIDS and allergies), infections
CC (e.g. bacterial and viral), metabolic disorders (e.g. obesity),
CC reproductive disorders (e.g. infertility), neurological disorders (e.g.
CC Parkinson's disease and Alzheimer's disease), cardiovascular disorders
CC (e.g. myocardial infarction and hypertension), eye disorders, or cell
CC proliferative diseases (e.g. cancer). The present DNA sequence encodes a
CC human NAAP protein of the invention.
XX
SQ Sequence 1360 BP; 464 A; 265 C; 325 G; 306 T; 0 U; 0 Other;
Query Match 56.1%; Score 23; DB 12; Length 1360;
Best Local Similarity 74.4%; Pred. No. 34;
Matches 29; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
QY 1 TCTCTCTCGACTTTGTCGTAATTTGATAAGATCTAATT 39
|||||
DB 561 TCTCTCATGTCTTCGATAGGAATTTATAGGTCCTAATT 523
|||||
RESULT 15
ADB62733/C
ID ADB62733 standard; cDNA; 2069 BP.
XX
XX ADB62733;
XX AC
XX 04-DEC-2003 (first entry)
XX
DE Human cDNA encoding clone NT2NE20060750.
XX
XX Human; ss; gene; pharmaceutical; diagnostic; gene therapy;
KW tissue regeneration; cell regeneration; membrane protein;
KW signal transduction-related protein; transcription-related protein;
KW osteoporosis; neurological disease; cancer; tumour.
XX

OS Homo sapiens.
XX
XX Key Location/Qualifiers
FT CDS 239..328
FT /*tag= a
FT /product= "Clone NT2NE20060750 protein"
XX
XX EPI308459-A2.
XX PN
XX PD 07-MAY-2003.
XX
XX 28-MAR-2002; 2002EP-00007401.
XX
XX 05-NOV-2001; 2001JP-00379298.
PR 25-JAN-2002; 2002US-00350978.
XX
XX (HELI-) HELIX RES INST.
PA (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
XX Iisogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;
XX
DR WPI; 2003-450961/43.
DR P-PSDB; ADB64703.
XX
XX New polynucleotides and polypeptides, useful for developing a diagnostic
PT marker or medicines for regulation of their expression and activity, or
PT as targets of gene therapy.
XX
PS Claim 1; Page; 222pp; English.
XX
XX The invention discloses a polynucleotide comprising a sequence selected
CC from 1970 fully defined nucleotide sequences which encode novel
CC polypeptides. Also claimed is a polypeptide encoded by the polynucleotide
CC or its partial peptide, an antibody binding to the polypeptide or peptide
CC of the polynucleotide, immunologically assaying the polypeptide or
CC peptide of the polynucleotide by contacting the polypeptide or peptide
CC with the antibody of the encoded protein, and observing the binding
CC between the two, a transformant carrying the polynucleotide in an
CC expressible manner and an antisense polynucleotide. The oligonucleotide
CC is useful as a primer for synthesising the polynucleotide, or as a probe
CC for detecting the polynucleotide. The polynucleotides and encoded
CC proteins are useful as pharmaceutical agents and many disease-related
CC genes may be included in them, for developing a diagnostic marker or
CC medicines for regulation of their expression and activity, or as targets
CC of gene therapy. The genes are involved in tissue and/or cell
CC regeneration. Membrane proteins, signal transduction-related proteins,
CC encoding them can be used as indicators for diseases (e.g. osteoporosis,
CC neurological diseases, cancer, tumours). The cDNA may be used to regulate
CC the activity or expression of the encoded protein to treat diseases. The
CC sequence presented is a cDNA of the invention. Note: Some of the sequence
CC data for this patent is not represented in the printed specification, but
CC is based on sequence information supplied by the European Patent Office.
XX
SQ Sequence 2069 BP; 650 A; 425 C; 504 G; 490 T; 0 U; 0 Other;
Query Match 56.1%; Score 23; DB 10; Length 2069;
Best Local Similarity 74.4%; Pred. No. 37;
Matches 29; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
QY 1 TCTCTCTCGACTTTGTCGTAATTTGATAAGATCTAATT 39
|||||
DB 855 TCTCTCATGTCTTCGATAGGAATTTATAGGTCCTAATT 817
|||||

Search completed: November 3, 2004, 06:59:24
Job time : 88.8636 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 3, 2004, 06:51:04 ; Search time 74.6435 Seconds
(without alignments)
2816.475 Million cell updates/sec

Title: US-09-933-267A-1_COPY_64311_64351

Perfect score: 41

Sequence: 1 tctctctgacttgcgtg.....aattgataagatctaatttg 41

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3413475 seqs, 2563800928 residues

Total number of hits satisfying chosen parameters: 6826950

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Published Applications NA:*

- 1: /cgn2_6/prodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/prodata/2/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/prodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/prodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/prodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/prodata/2/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/prodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/prodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/prodata/2/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/prodata/2/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/prodata/2/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/prodata/2/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/prodata/2/pubpna/US10A_PUBCOMB.seq:*
- 14: /cgn2_6/prodata/2/pubpna/US10B_PUBCOMB.seq:*
- 15: /cgn2_6/prodata/2/pubpna/US10C_PUBCOMB.seq:*
- 16: /cgn2_6/prodata/2/pubpna/US10D_PUBCOMB.seq:*
- 17: /cgn2_6/prodata/2/pubpna/US10E_PUBCOMB.seq:*
- 18: /cgn2_6/prodata/2/pubpna/US10_NEW_PUB.seq:*
- 19: /cgn2_6/prodata/2/pubpna/US11_NEW_PUB.seq:*
- 20: /cgn2_6/prodata/2/pubpna/US60_NEW_PUB.seq:*
- 21: /cgn2_6/prodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	41	100.0	465237	9	US-09-933-267A-1
2	39.4	96.1	392000	15	US-10-027-983-11
3	39.4	96.1	392000	15	US-10-448-753-11
4	24	58.5	149062	17	US-10-367-094-93
5	23.8	58.0	1368	9	US-09-815-242-3877
6	23.8	58.0	1389	16	US-10-282-122A-6247
7	23.8	58.0	1392	9	US-09-815-242-6565
8	23.4	57.1	591	9	US-09-864-761-14894
9	23	56.1	2069	15	US-10-104-047-887
10	23	56.1	78072	15	US-10-085-117-154
11	22.4	54.6	2709	16	US-10-282-122A-39149
12	22	53.7	5865	15	US-10-369-493-46473

13	53.7	32172	10	US-09-764-891-6161	Sequence 6161, Ap
14	53.2	434	16	US-10-424-593-23624	Sequence 23624, A
15	53.2	2047	16	US-10-425-114-12000	Sequence 12000, A
16	53.2	2070	16	US-10-424-593-121294	Sequence 121294, A
17	53.2	2621	15	US-10-104-047-949	Sequence 949, App
18	52.7	982	16	US-10-424-593-100224	Sequence 100224, A
19	52.7	2024	9	US-09-887-576-242	Sequence 242, App
20	52.2	1541	16	US-10-282-122A-36911	Sequence 36911, A
21	52.2	3225	17	US-10-437-963-24292	Sequence 24292, A
22	51.7	474	18	US-10-674-124A-11922	Sequence 11922, A
23	51.7	824	16	US-10-424-593-41879	Sequence 41879, A
24	51.2	437	13	US-10-027-633-182271	Sequence 182271, A
25	51.2	437	15	US-10-027-633-182271	Sequence 182271, A
26	51.2	552	13	US-10-027-633-282130	Sequence 282130, A
27	51.2	552	13	US-10-027-633-282131	Sequence 282131, A
28	51.2	552	13	US-10-027-633-282132	Sequence 282132, A
29	51.2	552	13	US-10-027-633-282133	Sequence 282133, A
30	51.2	552	15	US-10-027-633-282130	Sequence 282130, A
31	51.2	552	15	US-10-027-633-282131	Sequence 282131, A
32	51.2	552	15	US-10-027-633-282132	Sequence 282132, A
33	51.2	552	15	US-10-027-633-282133	Sequence 282133, A
34	51.2	819	15	US-10-032-585-6455	Sequence 6455, Ap
35	51.2	3664	16	US-10-108-260A-35	Sequence 35, Appl
36	51.2	9540	17	US-10-325-468-46	Sequence 46, Appl
37	51.2	10872	16	US-10-221-613-250	Sequence 250, App
38	50.7	455	17	US-10-437-963-72997	Sequence 72997, A
39	50.7	2238	15	US-10-032-585-6125	Sequence 6125, A
40	50.7	2370	16	US-10-282-122A-18027	Sequence 18027, A
41	50.7	2698	16	US-10-469-193-5	Sequence 5, Appl
42	50.7	2709	16	US-10-282-122A-39991	Sequence 39991, A
43	50.7	8339	8	US-08-781-986A-175	Sequence 175, App
44	50.7	8339	16	US-10-329-624-175	Sequence 175, App
45	50.7	13069	9	US-09-764-869-1850	Sequence 1850, Ap

ALIGNMENTS

RESULT 1
US-09-933-267A-1
; Sequence 1, Application US/09933267A
; Patent No. US20020123095A1
; GENERAL INFORMATION:
; APPLICANT: Kalush, Francis et al.
; TITLE OF INVENTION: Estrogen receptor alpha variants and
; TITLE OF INVENTION: methods of detection thereof
; FILE REFERENCE: CL000258C14
; CURRENT APPLICATION NUMBER: US/09/933,267A
; PRIOR FILING DATE: 2001-08-21
; PRIOR APPLICATION NUMBER: 60/160626
; PRIOR FILING DATE: 1999-10-20
; PRIOR APPLICATION NUMBER: 60/183756
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: 09/692414
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 09/768184
; PRIOR FILING DATE: 2001-01-24
; PRIOR APPLICATION NUMBER: 09/804076
; PRIOR FILING DATE: 2001-03-13
; PRIOR APPLICATION NUMBER: 09/826314
; PRIOR FILING DATE: 2001-04-05
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 465237
; TYPE: DNA
; ORGANISM: human
US-09-933-267A-1

Query Match 100.0%; Score 41; DB 9; Length 465237;
Best Local Similarity 100.0%; Pred. No. 2.8e-05;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy 1 TCTCTCGTACTTTCGTAATTAATGATAAGATCTAAATTTG 41
    |||||||
Db 64311 TCTCTCGTACTTTCGTAATTAATGATAAGATCTAAATTTG 64351

RESULT 2
US-10-027-983-11
; Sequence 11, Application US/10027983
; Publication No. US20030139360A1
; GENERAL INFORMATION:
; APPLICANT: Kenneth W. Dobie
; APPLICANT: Mark P. Roach
; TITLE OF INVENTION: ANTISENSE MODULATION OF ESTROGEN RECEPTOR ALPHA EXPRESSION
; FILE REFERENCE: RTS-0340
; CURRENT APPLICATION NUMBER: US/10/027,983
; CURRENT FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: US/10/027,983
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 11
; LENGTH: 392000
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 137740
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 137742
; OTHER INFORMATION: unknown
; NAME/KEY: misc feature
; LOCATION: (138122)...(139221)
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: unsure
; LOCATION: 145507
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 151967
; OTHER INFORMATION: unknown
; NAME/KEY: misc feature
; LOCATION: (151967)...(1542066)
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: unsure
; LOCATION: 154217
; OTHER INFORMATION: unknown
; NAME/KEY: misc feature
; LOCATION: (164037)...(164136)
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: misc feature
; LOCATION: (174657)...(174756)
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: misc feature
; LOCATION: (186224)...(186323)
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: misc feature
; LOCATION: (195242)...(195341)
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: unsure
; LOCATION: 202703
; OTHER INFORMATION: unknown
; NAME/KEY: misc feature
; LOCATION: (202771)...(202870)
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: misc feature
; LOCATION: (206246)...(215602)
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: misc feature
; LOCATION: (218126)...(218225)
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: misc feature
; LOCATION: (220360)...(220459)
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: misc feature
; LOCATION: (222717)...(222816)
; OTHER INFORMATION: n = A,T,C or G

```

```

; NAME/KEY: misc feature
; LOCATION: (223981)...(224080)
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: misc feature
; LOCATION: (227487)...(227586)
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: misc feature
; LOCATION: (230157)...(230256)
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: misc feature
; LOCATION: (232259)...(232398)
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: misc feature
; LOCATION: (236552)...(2366651)
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: misc feature
; LOCATION: (238789)...(248788)
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: exon
; LOCATION: (118288)...(119101)
; OTHER INFORMATION: exon 1C
; NAME/KEY: exon:intron junction
; LOCATION: (151129)...(151130)
; OTHER INFORMATION: exon 5:intron 5
; NAME/KEY: exon:intron junction
; LOCATION: (299248)...(299249)
; OTHER INFORMATION: exon 9:intron 9
; NAME/KEY: exon:intron junction
; LOCATION: (348578)...(348579)
; OTHER INFORMATION: exon 10:intron 10
; NAME/KEY: intron
; LOCATION: (348579)...(381838)
; OTHER INFORMATION: intron 10
; NAME/KEY: intron:exon junction
; LOCATION: (386185)...(386186)
; OTHER INFORMATION: intron 11:exon 12
US-10-027-983-11

Query Match 96.1%; Score 39,4; DB 15; Length 392000;
Best Local Similarity 97.6%; Pred No. 0.00011;
Matches 40; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TCTCTCGTACTTTCGTAATTAATGATAAGATCTAAATTTG 41
    |||||||
Db 12773 TCTCTCGTACTTTCGTAATTAATGATAAGATCTAAATTTG 12813

RESULT 3
US-10-448-753-11
; Sequence 11, Application US/10448753
; Publication No. US20030211611A1
; GENERAL INFORMATION:
; APPLICANT: Kenneth W. Dobie
; APPLICANT: Mark P. Roach
; TITLE OF INVENTION: ANTISENSE MODULATION OF ESTROGEN RECEPTOR ALPHA EXPRESSION
; FILE REFERENCE: RTS-0340
; CURRENT APPLICATION NUMBER: US/10/448,753
; CURRENT FILING DATE: 2003-05-30
; PRIOR APPLICATION NUMBER: US/10/027,983
; PRIOR FILING DATE: 2001-12-18
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 11
; LENGTH: 392000
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 137740
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 137742
; OTHER INFORMATION: unknown

```

```
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (138122)...(138221)
; OTHER INFORMATION: n = A,T,C or G
;
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 145507
; OTHER INFORMATION: unknown
;
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 151967
; OTHER INFORMATION: unknown
;
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (151967)...(1542066)
; OTHER INFORMATION: n = A,T,C or G
;
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 154217
; OTHER INFORMATION: unknown
;
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (164037)...(164136)
; OTHER INFORMATION: n = A,T,C or G
;
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (174657)...(174756)
; OTHER INFORMATION: n = A,T,C or G
;
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (186224)...(186323)
; OTHER INFORMATION: n = A,T,C or G
;
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (195242)...(195341)
; OTHER INFORMATION: n = A,T,C or G
;
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 202703
; OTHER INFORMATION: unknown
;
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (202771)...(202870)
; OTHER INFORMATION: n = A,T,C or G
;
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (206246)...(215602)
; OTHER INFORMATION: n = A,T,C or G
;
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (218126)...(218225)
; OTHER INFORMATION: n = A,T,C or G
;
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (220360)...(220459)
; OTHER INFORMATION: n = A,T,C or G
;
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (222717)...(222816)
; OTHER INFORMATION: n = A,T,C or G
;
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (223981)...(224080)
; OTHER INFORMATION: n = A,T,C or G
;
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (227487)...(227586)
; OTHER INFORMATION: n = A,T,C or G
;
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (230157)...(230256)
; OTHER INFORMATION: n = A,T,C or G
;
; FEATURE:
```

```
; NAME/KEY: misc feature
; LOCATION: (232299)...(232398)
; OTHER INFORMATION: n = A,T,C or G
;
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (236552)...(2366651)
; OTHER INFORMATION: n = A,T,C or G
;
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (238789)...(248788)
; OTHER INFORMATION: n = A,T,C or G
;
; FEATURE:
; NAME/KEY: exon
; LOCATION: (118288)...(119101)
; OTHER INFORMATION: exon 1C
;
; FEATURE:
; NAME/KEY: exon:intron junction
; LOCATION: (151129)...(151130)
; OTHER INFORMATION: exon 5:intron 5
;
; FEATURE:
; NAME/KEY: exon:intron junction
; LOCATION: (299248)...(299249)
; OTHER INFORMATION: exon 9:intron 9
;
; FEATURE:
; NAME/KEY: exon:intron junction
; LOCATION: (348578)...(348579)
; OTHER INFORMATION: exon 10:intron 10
;
; FEATURE:
; NAME/KEY: intron
; LOCATION: (348579)...(381838)
; OTHER INFORMATION: intron 10
;
; FEATURE:
; NAME/KEY: intron:exon junction
; LOCATION: (386185)...(386186)
; OTHER INFORMATION: intron 11:exon 12
;
; US-10-448-753-11
```

Query Match 96.1%; Score 39.4; DB 15; Length 392000;
Best Local Similarity 97.6%; Pred. No. 0.00011;
Matches 40; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TCCTCCTGACTTTGTCGTAATTTGATAAGATCTAAATTTG 41
|||
Db 12773 TCCTCCTGACTTTGTCGTAATTTGATAAGATCTAAATTTG 12813

RESULT 4
US-10-367-094-93
; Sequence 93, Application US/10367094
; Publication No. US20040170982A1
; GENERAL INFORMATION:
; APPLICANT: David W. Morris
; APPLICANT: Marc Malandro
; TITLE OF INVENTION: Novel
; FILE REFERENCE: 529452001500
; CURRENT APPLICATION NUMBER: US/10/367,094
; CURRENT FILING DATE: 2003-02-14
; NUMBER OF SEQ ID NOS: 203
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 93
; LENGTH: 149062
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-367-094-93

Query Match 58.5%; Score 24; DB 17; Length 149062;
Best Local Similarity 75.0%; Pred. No. 93;
Matches 30; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 1 TCCTCCTGACTTTGTCGTAATTTGATAAGATCTAAATTT 40
|||
Db 124343 TCACTGATACTTACTCAGTGAATATATAAGATCTAAATTT 124382

```
RESULT 5
US-09-815-242-3877/c
; Sequence 3877, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3877
; LENGTH: 1368
; TYPE: DNA
; ORGANISM: Enterococcus faecalis
US-09-815-242-3877

Query Match 58.0%; Score 23.8; DB 9; Length 1368;
Best Local Similarity 80.0%; Pred. No. 35;
Matches 28; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 3 TCTCTGACTTTGTCGTAATTTGATAGATCTAA 37
||||| ||||||| ||||||| ||||||| |||
DB 1193 TCTCTTTCATTTGTCGTAATTTGATATTCCTAA 1159

RESULT 6
US-10-282-122A-6247/c
; Sequence 6247, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
```

```
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6247
; LENGTH: 1389
; TYPE: DNA
; ORGANISM: Enterococcus faecalis
US-10-282-122A-6247

Query Match 58.0%; Score 23.8; DB 16; Length 1389;
Best Local Similarity 80.0%; Pred. No. 35;
Matches 28; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 3 TCTCTGACTTTGTCGTAATTTGATAGATCTAA 37
||||| ||||||| ||||||| ||||||| |||
DB 1196 TCTCTTTCATTTGTCGTAATTTGATATTCCTAA 1162

RESULT 7
US-09-815-242-6565/c
; Sequence 6565, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6565
; LENGTH: 1392
; TYPE: DNA
```



```
; ORGANISM: Enterococcus faecalis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1392)
US-09-815-242-6365

Query Match      58.0%; Score 23.8; DB 9; Length 1392;
Best Local Similarity 80.0%; Pred. No. 35;
Matches 28; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 3 TCTCCTGACTTTGTCGTAATGATAAGATCTAA 37
    ||||| ||||| ||||| ||||| ||||| |||||
Db 1196 TCTCCTTTCATGTCGTAATGATAAGATCTAA 1162

RESULT 8
US-09-864-761-14894
; Sequence 14894, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeonica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
; SEQ ID NO 14894
; LENGTH: 591
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC007719.1
```

```
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.8
US-09-864-761-14894

Query Match      57.1%; Score 23.4; DB 9; Length 591;
Best Local Similarity 73.2%; Pred. No. 40;
Matches 30; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 1 TCTCCTGACTTTGTCGTAATGATAAGATCTAAATTG 41
    ||||| ||||| ||||| ||||| ||||| |||||
Db 80 TCTCTACTGACTTCTCTTTAATTAGCTTAGTGTAAATTG 120

RESULT 9
US-10-104-047-887/c
; Sequence 887, Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20030236392A1el full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 887
; LENGTH: 2069
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-104-047-887

Query Match      56.1%; Score 23; DB 15; Length 2069;
Best Local Similarity 74.4%; Pred. No. 79;
Matches 29; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1 TCTCCTGACTTTGTCGTAATGATAAGATCTAAATT 39
    ||||| ||||| ||||| ||||| ||||| |||||
Db 855 TCTCTCATGCTTCGATAGGAAATTTATAAGGCTATT 817

RESULT 10
US-10-085-117-154/c
; Sequence 154, Application US/10085117
; Publication No. US2003023334A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
; FILE REFERENCE: 529452000121
; CURRENT APPLICATION NUMBER: US/10/085,117
; CURRENT FILING DATE: 2002-02-27
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 361
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 154
; LENGTH: 78072
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-085-117-154

Query Match      56.1%; Score 23; DB 15; Length 78072;
Best Local Similarity 74.4%; Pred. No. 2e+02;
Matches 29; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 2 CTCTCCTGACTTTGTCGTAATGATAAGATCTAAATT 40
    ||||| ||||| ||||| ||||| ||||| |||||
Db 42822 CTCTCCTGACTTAGTTGTAATCGAGACAGATGTAAC 42784
```

```
RESULT 11
US-10-282-122A-39149
; Sequence 39149, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 39149
; LENGTH: 2709
; TYPE: DNA
; ORGANISM: Salmonella typhimurium
US-10-282-122A-39149
Query Match 54.6%; Score 22.4; DB 16; Length 2709;
Best Local Similarity 72.5%; Pred. No. 1.4e+02;
Matches 29; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 2 CTCCTCGACTTTGTCGCGTAAATTGATAGATCTAAATTG 41
|||||
Db 1309 CTCCTCGACGCTGCGACAGAAAATCGATGAGATCCGGTTG 1348

RESULT 12
US-10-369-493-46473
; Sequence 46473, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
US-10-369-493-46473
Query Match 54.6%; Score 22.4; DB 16; Length 2709;
Best Local Similarity 72.5%; Pred. No. 1.4e+02;
Matches 29; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 2 CTCCTCGACTTTGTCGCGTAAATTGATAGATCTAAATTG 41
|||||
Db 1309 CTCCTCGACGCTGCGACAGAAAATCGATGAGATCCGGTTG 1348

RESULT 13
US-09-764-891-6161
; Sequence 6161, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6161
; LENGTH: 32172
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-6161
Query Match 53.7%; Score 22; DB 10; Length 32172;
Best Local Similarity 73.7%; Pred. No. 3.9e+02;
Matches 28; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 3 TCTCTCGACTTTGTCGCGTAAATTGATAGATCTAAATT 40
|||||
Db 25309 TCTACTGACTCTCTCATTAATTAATGAATGATCTCT 25346

RESULT 14
US-10-424-599-23624/c
; Sequence 23624, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 23624
; LENGTH: 434
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_121333C.1
US-10-424-599-23624
Query Match 53.2%; Score 21.8; DB 16; Length 434;
```

Best Local Similarity 78.8%; Pred. No. 1.6e+02;
Matches 26; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
QY 7 CTGACTTTGTCGTAATTCATAGATCTTAATT 39
Db 186 CTGCCGCTGTCATAATAATTTATAAGATTTAATT 154

RESULT 15
US-10-425-114-12000/c
; Sequence 12000, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 12000
; LENGTH: 2047
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: 701152196_FLI
US-10-425-114-12000

Query Match 53.2%; Score 21.8; DB 16; Length 2047;
Best Local Similarity 78.8%; Pred. No. 2.3e+02;
Matches 26; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
QY 9 GACTTTGTCGTAATTCATAGATCTTAATTG 41
Db 1172 GACTTTGACGGGAATGATGATGATTTAGTTG 1140

Search completed: November 3, 2004, 10:01:03
Job time : 77.6435 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 3, 2004, 05:58:29 ; Search time 18.8325 Seconds
(without alignments)
1547.447 Million cell updates/sec

Title: US-09-933-267A-1_COPY_64311_64351

Perfect score: 41

Sequence: 1 tctctcgtgactttgtccgt.....aattgataagatctaattg 41

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 824507 seqs, 355394441 residues

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:*

- 1: /cgn2_6/prodata/1/ina/5A COMB.seq.*
- 2: /cgn2_6/prodata/1/ina/5B COMB.seq.*
- 3: /cgn2_6/prodata/1/ina/6A COMB.seq.*
- 4: /cgn2_6/prodata/1/ina/6B COMB.seq.*
- 5: /cgn2_6/prodata/1/ina/PCRUS COMB.seq.*
- 6: /cgn2_6/prodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39.4	96.1	392000	4	US-10-027-983-11
2	23.8	58.0	1347	4	US-09-134-000C-389
3	21.4	52.2	497	4	US-09-270-767-6174
4	21.4	52.2	497	4	US-09-270-767-21456
5	21.2	51.7	395	2	US-08-967-101-19
6	21.2	51.7	395	2	US-08-992-541-19
7	21.2	51.7	395	3	US-09-124-698-19
8	21.2	51.7	395	3	US-09-127-480-19
9	21.2	51.7	395	3	US-08-496-841C-19
10	21.2	51.7	395	3	US-09-124-523-19
11	21.2	51.7	395	4	US-09-636-796A-19
12	21.2	51.7	395	4	US-08-431-048F-19
13	21.2	51.2	465	4	US-09-248-796A-5348
14	20.8	50.7	237	4	US-09-248-796A-1647
15	20.8	50.7	951	4	US-09-543-681A-1701
16	20.8	50.7	1557	3	US-07-956-483-25
17	20.8	50.7	1557	3	US-08-472-240A-17
18	20.8	50.7	2238	4	US-09-792-024-34
19	20.8	50.7	8339	4	US-08-956-171B-175
20	20.8	50.7	8339	4	US-08-781-986A-175
21	20.6	50.2	379	4	US-09-513-999C-15941
22	20.6	50.2	8549	4	US-08-956-171B-5
23	20.6	50.2	8549	4	US-08-781-986A-5
24	20.4	49.8	389	4	US-09-702-705-1766
25	20.4	49.8	389	4	US-09-736-457-1766
26	20.4	49.8	389	4	US-09-671-325-1766
27	20.4	49.8	389	4	US-09-658-824-1766

C 28	20.4	49.8	1590	3	US-09-334-938-8	Sequence 8, Appli
C 29	20.4	49.8	2365	4	US-08-956-171B-95	Sequence 95, Appl
C 30	20.4	49.8	2365	4	US-08-781-986A-95	Sequence 95, Appl
C 31	20.2	49.3	494	4	US-09-621-976-1823	Sequence 1823, Ap
C 32	20.2	49.3	1032	4	US-09-543-681A-2287	Sequence 2287, Ap
C 33	20.2	49.3	2904	3	US-09-221-294-3	Sequence 3, Appli
C 34	20.2	49.3	3428	4	US-09-919-039-77	Sequence 77, Appl
C 35	20	48.8	201	4	US-09-248-796A-8344	Sequence 8344, Ap
C 36	20	48.8	522	4	US-09-248-796A-6933	Sequence 6933, Ap
C 37	20	48.8	603	4	US-09-248-796A-13707	Sequence 13707, A
C 38	20	48.8	849	4	US-09-248-796A-435	Sequence 435, App
C 39	19.8	48.3	265	3	US-08-905-223-102	Sequence 102, App
C 40	19.8	48.3	797	3	US-09-071-353-1	Sequence 1, Appli
C 41	19.8	48.3	797	4	US-09-426-326-1	Sequence 1, Appli
C 42	19.8	48.3	1320	4	US-09-248-796A-5717	Sequence 5717, Ap
C 43	19.6	47.8	588	4	US-09-220-132-101	Sequence 101, App
C 44	19.4	47.3	204	4	US-09-583-110-2281	Sequence 2281, Ap
C 45	19.4	47.3	204	4	US-09-583-110-2284	Sequence 2284, Ap

ALIGNMENTS

RESULT 1
US-10-027-983-11
; Sequence 11, Application US/10027983
; Patent No. 6617162
; GENERAL INFORMATION:
; APPLICANT: Kenneth W. Dobie
; TITLE OF INVENTION: ANTISENSE MODULATION OF ESTROGEN RECEPTOR ALPHA EXPRESSION
; FILE REFERENCE: RTS-0340
; CURRENT APPLICATION NUMBER: US-10-027,983
; CURRENT FILING DATE: 2001-12-18
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 11
; LENGTH: 392000
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 137740
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 137742
; OTHER INFORMATION: unknown
; NAME/KEY: misc feature
; LOCATION: (138122)...(138221)
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: unsure
; LOCATION: 145507
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 151967
; OTHER INFORMATION: unknown
; NAME/KEY: misc feature
; LOCATION: (151967)...(1542066)
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: unsure
; LOCATION: 154217
; OTHER INFORMATION: unknown
; NAME/KEY: misc feature
; LOCATION: (154037)...(164136)
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: misc feature
; LOCATION: (174657)...(174756)
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: misc feature
; LOCATION: (186224)...(186323)
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: misc feature
; LOCATION: (195242)...(195341)
; OTHER INFORMATION: n = A,T,C or G

Search 5

no 2003?

#1

NAME/KEY: unsure
LOCATION: 202703
OTHER INFORMATION: unknown
NAME/KEY: misc feature
LOCATION: (202711)...(202870)
OTHER INFORMATION: n = A,T,C or G
NAME/KEY: misc feature
LOCATION: (206246)...(215602)
OTHER INFORMATION: n = A,T,C or G
NAME/KEY: misc feature
LOCATION: (218126)...(219225)
OTHER INFORMATION: n = A,T,C or G
NAME/KEY: misc feature
LOCATION: (220360)...(220459)
OTHER INFORMATION: n = A,T,C or G
NAME/KEY: misc feature
LOCATION: (222717)...(222816)
OTHER INFORMATION: n = A,T,C or G
NAME/KEY: misc feature
LOCATION: (223981)...(224080)
OTHER INFORMATION: n = A,T,C or G
NAME/KEY: misc feature
LOCATION: (227487)...(227586)
OTHER INFORMATION: n = A,T,C or G
NAME/KEY: misc feature
LOCATION: (230157)...(230256)
OTHER INFORMATION: n = A,T,C or G
NAME/KEY: misc feature
LOCATION: (232299)...(232398)
OTHER INFORMATION: n = A,T,C or G
NAME/KEY: misc feature
LOCATION: (236552)...(2366651)
OTHER INFORMATION: n = A,T,C or G
NAME/KEY: misc feature
LOCATION: (238789)...(248788)
OTHER INFORMATION: n = A,T,C or G
NAME/KEY: exon
LOCATION: (118288)...(119101)
OTHER INFORMATION: exon 1C
NAME/KEY: exon:intron junction
LOCATION: (151129)...(151130)
OTHER INFORMATION: exon 5:intron 5
NAME/KEY: exon:intron junction
LOCATION: (299248)...(299249)
OTHER INFORMATION: exon 9:intron 9
NAME/KEY: exon:intron junction
LOCATION: (348578)...(348579)
OTHER INFORMATION: exon 10:intron 10
NAME/KEY: intron
LOCATION: (348579)...(381838)
OTHER INFORMATION: intron 10
NAME/KEY: intron:exon junction
LOCATION: (386185)...(386186)
OTHER INFORMATION: intron 11:exon 12
US-10-027-983-11

Query Match 96.1%; Score 39.4; DB 4; Length 392000;
Best Local Similarity 97.6%; Pred. No. 1.8e-06;
Matches 40; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TCTCTCTGACTTTGTCGTAATTTGATAAGATCTAAATTG 41
Db 12773 TCTCTCTGACTTTGTCGTAATTTGATAAGATCTAAATTG 12813

RESULT 2

US-09-134-000C-389/c
Sequence 389, Application US/09134000C
Patent No. 6617156
GENERAL INFORMATION:

APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 032796-032
CURRENT APPLICATION NUMBER: US/09/134,000C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/055,778
PRIOR FILING DATE: 1997-08-15
NUMBER OF SEQ ID NOS: 6812
SOFTWARE: PatentIn version 3.1
SEQ ID NO 389
LENGTH: 1347
TYPE: DNA
ORGANISM: Enterococcus faecalis
US-09-134-000C-389

Query Match 58.0%; Score 23.8; DB 4; Length 1347;
Best Local Similarity 80.0%; Pred. No. 1.3;
Matches 28; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 3 TCTCTCTGACTTTGTCGTAATTTGATAAGATCTAA 37
Db 1151 TCTCTCTGACTTTGTCGTAATTTGATAATTCCTAA 1117

RESULT 3

US-09-270-767-6174
Sequence 6174, Application US/09270767
Patent No. 6703491
GENERAL INFORMATION:

APPLICANT: Homburger et al.
TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*
FILE REFERENCE: File Reference: 7326-094
CURRENT APPLICATION NUMBER: US/09/270,767
CURRENT FILING DATE: 1999-03-17
NUMBER OF SEQ ID NOS: 62517
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 6174
LENGTH: 497
TYPE: DNA
ORGANISM: *Drosophila melanogaster*
US-09-270-767-6174

Query Match 52.2%; Score 21.4; DB 4; Length 497;
Best Local Similarity 71.8%; Pred. No. 10;
Matches 28; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 1 TCTCTCTGACTTTGTCGTAATTTGATAAGATCTAAATT 39
Db 330 TCTCTCTGACTTTGTCGTAATTTGATAAGATCTAAATT 368

RESULT 4

US-09-270-767-21456
Sequence 21456, Application US/09270767
Patent No. 6703491
GENERAL INFORMATION:

APPLICANT: Homburger et al.
TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*
FILE REFERENCE: File Reference: 7326-094
CURRENT APPLICATION NUMBER: US/09/270,767
CURRENT FILING DATE: 1999-03-17
NUMBER OF SEQ ID NOS: 62517
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 21456
LENGTH: 497
TYPE: DNA
ORGANISM: *Drosophila melanogaster*
US-09-270-767-21456

Query Match 52.2%; Score 21.4; DB 4; Length 497;
Best Local Similarity 71.8%; Pred. No. 10;
Matches 28; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 1 TCTCTCTGACTTTGTCGTAATTTGATAAGATCTAAATT 39
Db 1151 TCTCTCTGACTTTGTCGTAATTTGATAATTCCTAA 1117

```

; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/592,541
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Pitcher, Edmund R.
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 395 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-592-541-19

Query Match 51.7%; Score 21.2; DB 2; Length 395;
Best Local Similarity 76.5%; Pred. No. 11;
Matches 26; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 5 TCTGCACTTGTCCGTAATTTGATAAGATCTAAT 38
| | | | | | | | | | | | | | | | | | | |
Db 271 TCTGAGTTTCTCCATAAATTAATGACCTAAT 304

RESULT 7
US-09-124-698-19
; Sequence 19, Application US/09124698
; Patent No. 6117978
; GENERAL INFORMATION:
; APPLICANT: ST. GEORGE-HYSLOP, PETER H
; APPLICANT: ROMMENS, JOHANNA M
; APPLICANT: FRASER, PAUL E
; TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
; TITLE OF INVENTION: TO ALZHEIMER'S DISEASE
; NUMBER OF SEQUENCES: 183
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
; STREET: High Street tower - 125 High Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/124,698
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/592,541
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Pitcher, Edmund R.
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 395 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single

```

```
;
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-09-124-698-19

Query Match      51.7%; Score 21.2; DB 3; Length 395;
Best Local Similarity 76.5%; Pred. No. 11;
Matches 26; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 5 TCCTGACTTTCGCTAAATGATAAGATCTAAT 38
Db 271 TTCTGAGTTTCTCCATAAATTAATTGGACCTAAT 304

RESULT 8
US-09-127-480-19
; Sequence 19, Application US/09127480
; Patent No. 6194153
; GENERAL INFORMATION:
; APPLICANT: ST. GEORGE-HYSLOP, PETER H
; APPLICANT: ROMMENS, JOHANNA M
; APPLICANT: FRASER, PAUL E
; TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
; NUMBER OF SEQUENCES: 183
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
; STREET: High Street Tower - 125 High Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: US/09/127,480
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/592,541
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Pitcher, Edmund R.
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 395 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-09-127-480-19

Query Match      51.7%; Score 21.2; DB 3; Length 395;
Best Local Similarity 76.5%; Pred. No. 11;
Matches 26; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 5 TCCTGACTTTCGCTAAATGATAAGATCTAAT 38
Db 271 TTCTGAGTTTCTCCATAAATTAATTGGACCTAAT 304

RESULT 9
US-08-496-841C-19
; Sequence 19, Application US/08496841C
; Patent No. 6210919
; GENERAL INFORMATION:
; APPLICANT: ST. GEORGE-HYSLOP, PETER H
; APPLICANT: ROMMENS, JOHANNA M
```

```
;
; FRASER, PAUL E
; TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
; TO ALZHEIMER'S DISEASE
; NUMBER OF SEQUENCES: 175
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby, PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/496,841C
; FILING DATE: 28-Jun-1995
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Paul F. Fehlner, Ph.D.
; REGISTRATION NUMBER: 35,135
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 527-7700
; TELEFAX: (212) 753-6237
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 395 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; SEQUENCE DESCRIPTION: SEQ ID NO: 19:
US-08-496-841C-19

Query Match      51.7%; Score 21.2; DB 3; Length 395;
Best Local Similarity 76.5%; Pred. No. 11;
Matches 26; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 5 TCCTGACTTTCGCTAAATGATAAGATCTAAT 38
Db 271 TTCTGAGTTTCTCCATAAATTAATTGGACCTAAT 304

RESULT 10
US-09-124-523-19
; Sequence 19, Application US/09124523
; Patent No. 6395960
; GENERAL INFORMATION:
; APPLICANT: ST. GEORGE-HYSLOP, PETER H
; APPLICANT: ROMMENS, JOHANNA M
; APPLICANT: FRASER, PAUL E
; TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
; NUMBER OF SEQUENCES: 183
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
; STREET: High Street Tower - 125 High Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/124,523
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
```


APPLICATION NUMBER: US/08/592,541

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Pitcher, Edmund R.

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 248-7000

TELEFAX: (617) 248-7100

INFORMATION FOR SEQ ID NO: 19:

SEQUENCE CHARACTERISTICS:

LENGTH: 395 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-09-124-523-19

Query Match 51.7%; Score 21.2; DB 3; Length 395;

Best Local Similarity 76.5%; Pred. No. 11;

Matches 26; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 5 TCCTGACTTTCGCGTAAATTCGATGATCTAAT 38

DB 271 TTCTGAGTTTCTCCATAAATTAATGGACCTAAT 304

RESULT 11

US-09-636-796A-19

Sequence 19, Application US/09636796A

Patent No. 6485911

GENERAL INFORMATION:

APPLICANT: ST. GEORGE-HYSLOP, PETER H

ROMMENS, JOHANNA M

FRASER, PAUL E

TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED

TO ALZHEIMER'S DISEASE

NUMBER OF SEQUENCES: 183

CORRESPONDENCE ADDRESS:

ADDRESSEE: TESTA, HURWITZ & THIBEAULT

STREET: High Street Tower - 125 High Street

CITY: Boston

STATE: Massachusetts

COUNTRY: U.S.A.

ZIP: 02110

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/636,796A

FILING DATE: 11-Aug-2000

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/592,541

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Pitcher, Edmund R.

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 248-7000

TELEFAX: (617) 248-7100

INFORMATION FOR SEQ ID NO: 19:

SEQUENCE CHARACTERISTICS:

LENGTH: 395 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

SEQUENCE DESCRIPTION: SEQ ID NO: 19:

US-09-636-796A-19

Query Match 51.7%; Score 21.2; DB 4; Length 395;

Best Local Similarity 76.5%; Pred. No. 11;

Matches 26; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 5 TCCTGACTTTCGCGTAAATTCGATGATCTAAT 38

DB 271 TTCTGAGTTTCTCCATAAATTAATGGACCTAAT 304

RESULT 12

US-08-431-048F-19

Sequence 19, Application US/08431048F

Patent No. 6531586

GENERAL INFORMATION:

APPLICANT: ST. GEORGE-HYSLOP, PETER H

ROMMENS, JOHANNA M

FRASER, PAUL E

TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED

TO ALZHEIMER'S DISEASE

NUMBER OF SEQUENCES: 155

CORRESPONDENCE ADDRESS:

ADDRESSEE: DARBY & DARBY P.C.

STREET: 805 THIRD AVENUE

CITY: NEW YORK

STATE: N.Y.

COUNTRY: U.S.A.

ZIP: 10022-7513

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/431,048F

FILING DATE: 28-Apr-1995

CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: FEHLNER, PAUL F.

REGISTRATION NUMBER: 35135

REFERENCE/DOCKET NUMBER: 1034/0F808

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-527-7700

TELEFAX: 212-527-6237

INFORMATION FOR SEQ ID NO: 19:

SEQUENCE CHARACTERISTICS:

LENGTH: 395 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

SEQUENCE DESCRIPTION: SEQ ID NO: 19:

US-08-431-048F-19

Query Match 51.7%; Score 21.2; DB 4; Length 395;

Best Local Similarity 76.5%; Pred. No. 11;

Matches 26; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 5 TCCTGACTTTCGCGTAAATTCGATGATCTAAT 38

DB 271 TTCTGAGTTTCTCCATAAATTAATGGACCTAAT 304

RESULT 13

US-09-248-796A-5348

Sequence 5348, Application US/09248796A

Patent No. 6747137

GENERAL INFORMATION:

APPLICANT: Keith Weinstein et al

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS

FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 107196.132

CURRENT APPLICATION NUMBER: US/09/248,796A

CURRENT FILING DATE: 1999-02-12

PRIOR APPLICATION NUMBER: US 60/074,725

PRIOR FILING DATE: 1998-02-13

PRIOR APPLICATION NUMBER: US 60/096,409

; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 5348
; LENGTH: 465
; TYPE: DNA
; ORGANISM: Candida albicans
US-09-248-796A-5348

Query Match 51.2%; Score 21; DB 4; Length 465;
Best Local Similarity 73.0%; Pred. No. 14;
Matches 27; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 5 TCTGACTTTGTCGGTAAATGATAAGATCTAATTG 41
| | | | | | | | | | | | | | | | | | | | | |
Db 306 TACTGAGTGTCACTCAATTCATTGATTCATTAATTG 342

RESULT 14

US-09-248-796A-1647/c
; Sequence 1647, Application US/09248796A
; Patent No. 6747137

; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132

; CURRENT APPLICATION NUMBER: US/09/248,796A
; PRIOR FILING DATE: 1999-02-12

; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13

; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13

; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 1647

; LENGTH: 237
; TYPE: DNA

; ORGANISM: Candida albicans
US-09-248-796A-1647

Query Match 50.7%; Score 20.8; DB 4; Length 237;
Best Local Similarity 70.0%; Pred. No. 15;
Matches 28; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 1 TCTCTCTGACTTTGCCGTAATTCATAAGATCTAATTT 40
| | | | | | | | | | | | | | | | | | | | | |
Db 102 TTTATCTTTATTGTCGTTCAATTCCTAAAATATAATTT 63

RESULT 15

US-09-543-681A-1701/c
; Sequence 1701, Application US/09543681A
; Patent No. 6605709

; GENERAL INFORMATION:
; APPLICANT: GARY BRETON

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1002-001

; CURRENT APPLICATION NUMBER: US/09/543,681A
; PRIOR FILING DATE: 2000-04-05

; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09

; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 1701

; LENGTH: 951
; TYPE: DNA

; ORGANISM: Proteus mirabilis
US-09-543-681A-1701

Query Match 50.7%; Score 20.8; DB 4; Length 951;
Best Local Similarity 70.0%; Pred. No. 20;
Matches 28; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 1 TCTCTCTGACTTTGCCGTAATTCATAAGATCTAATTT 40

Db 46 TCACTCCTGATGTTTCAGTTAAATTCATATTAGCTAATT 7
| | | | | | | | | | | | | | | | | | | | | |

Search completed: November 3, 2004, 09:46:50
Job time : 19.8325 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 3, 2004, 04:55:34 ; Search time 255.809 Seconds
(without alignments)
7579.407 Million cell updates/sec

Title: US-09-933-267A-1_COPY_306362_306402

Perfect score: 41

Sequence: 1 ttgttgatgctgagccccc.....atactattccagatgatga 41

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4526729 seqs, 23644849745 residues

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb.ba.*

2: gb.htg.*

3: gb.in.*

4: gb.om.*

5: gb.ov.*

6: gb.pat.*

7: gb.ph.*

8: gb.pl.*

9: gb.pr.*

10: gb.ro.*

11: gb.sts.*

12: gb.sy.*

13: gb.un.*

14: gb.vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	41	100.0	392	9 AF120105	AF120105 Homo sapi
2	41	100.0	1218	6 AR397495	AR397495 Sequence
3	41	100.0	1218	9 AF172068	AF172068 Homo sapi
4	41	100.0	1218	9 AF172069	AF172069 Homo sapi
5	41	100.0	1223	6 AX066401	AX066401 Sequence
6	41	100.0	1237	9 HSESR13	AX066402 Sequence
7	41	100.0	1374	6 AX066402	AX066402 Sequence
8	41	100.0	1644	12 AF242866	AF242866 Synthetic
9	41	100.0	1770	6 AX824420	AX824420 Sequence
10	41	100.0	1788	6 AX411751	AX411751 Sequence
11	41	100.0	1803	6 AX824417	AX824417 Sequence
12	41	100.0	1956	6 AR116270	AR116270 Sequence
13	41	100.0	1983	6 AX411753	AX411753 Sequence
14	41	100.0	1983	6 AX411753	AX411753 Sequence
15	41	100.0	1983	6 AX411757	AX411757 Sequence
16	41	100.0	2092	6 I08538	I08538 Sequence 1
17	41	100.0	2092	6 AX474709	AX474709 Sequence
18	41	100.0	2092	9 HUMERMCF	M12674 Human estro
19	41	100.0	2106	9 HSU47678	U47678 Human 80 kD

20	41	100.0	2220	6 BD016996	BD016996 Fused est
21	41	100.0	2322	6 A42099	A42099 Sequence 1
22	41	100.0	2532	6 BD187708	BD187708 Specific
23	41	100.0	4963	6 AR029418	AR029418 Sequence
24	41	100.0	4963	6 I15368	I15368 Sequence 16
25	41	100.0	4963	6 I56760	I56760 Sequence 9
26	41	100.0	5439	9 HSM807087	AX640939 Homo sapi
27	41	100.0	6039	6 CO832274	CO832274 Sequence
28	41	100.0	6450	6 AR153585	AR153585 Sequence
29	41	100.0	6450	6 AR225684	AR225684 Sequence
30	41	100.0	6450	6 AR397400	AR397400 Sequence
31	41	100.0	6450	6 AX587952	AX587952 Sequence
32	41	100.0	6450	9 HSERR	X03635 Homo sapien
33	41	100.0	6567	6 AX128350	AX128350 Sequence
34	41	100.0	6610	12 AF061181	AF061181 Mammalian
35	41	100.0	6623	6 AX128345	AX128345 Sequence
36	41	100.0	6639	6 AX128351	AX128351 Sequence
37	41	100.0	6695	6 AX128347	AX128347 Sequence
38	41	100.0	6695	6 AX128353	AX128353 Sequence
39	41	100.0	6695	6 AX128354	AX128354 Sequence
40	41	100.0	6746	6 AX128344	AX128344 Sequence
41	41	100.0	6801	6 AX128352	AX128352 Sequence
42	41	100.0	6801	6 AX128355	AX128355 Sequence
43	41	100.0	6818	6 AX128346	AX128346 Sequence
44	41	100.0	6828	6 AX128340	AX128340 Sequence
45	41	100.0	6833	6 AX128349	AX128349 Sequence

ALIGNMENTS

AF120105 392 bp mRNA linear PRI 06-APR-1999
Homo sapiens alternatively-spliced estrogen receptor alpha mRNA,
partial cds.

AF120105 GI:4567039

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 392)

Campbell-Thompson, M.L.

Direct Submission

Submitted (13-JAN-1999) Medicine, University of Florida, Box 100214

Gastroenterology, Gainesville, FL 32657, USA

Location/Qualifiers

1..392

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/chromosome="6"

/map="6q25.1"

/sex="male"

/cell_type="colon adenocarcinoma"

<1..343

/note="steroid hormone receptor; missing exon 5"

/codon_start=2

/product="alternatively-spliced estrogen receptor alpha"

/protein_id="AAD23565.1"

/db_xref="GI:4567040"

/translation="RRGRMLKHKRQDDGGRGVGSAGDMRAANLWPSLMIKRSK

KNSLALSTADQMVSAALLDAEPPILYSEYDTPFPFSEASMMGLLNLADELVHINW

AKRVPCRTENV"

25

/note="compared to wild type sequence deposited in GenBank

Accession Number X03635"

/replace="a"

variation

ORIGIN

Query Match 100.0%; Score 41; DB 9; Length 392;

Best Local Similarity 100.0%; Pred. No. 8.5e-07;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TTGTGGATGCTGAGCCCCCATACTCTATTCCGAGTATGA 41
|||||
Db 182 TTGTGGATGCTGAGCCCCCATACTCTATTCCGAGTATGA 222

RESULT 2
LOCUS AR397495 1218 bp DNA linear PAT 18-DEC-2003
DEFINITION Sequence 98 from patent US 6617162.
ACCESSION AR397495
VERSION AR397495.1 GI:40134376

KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1218)
AUTHORS Dobie, K.W. and Roach, M.P.
TITLE Antisense modulation of estrogen receptor alpha expression
JOURNAL Patent: US 6617162-A 98 09-SEP-2003;
FEATURES Location/Qualifiers
source
1. .1218
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN

Query Match 100.0%; Score 41; DB 6; Length 1218;
Best Local Similarity 100.0%; Pred. No. 8.7e-07;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TTGTGGATGCTGAGCCCCCATACTCTATTCCGAGTATGA 41
|||||
Db 385 TTGTGGATGCTGAGCCCCCATACTCTATTCCGAGTATGA 425

RESULT 3
LOCUS AF172068 1218 bp mRNA linear PRI 02-OCT-2000
DEFINITION Homo sapiens fatty acid synthase/estrogen receptor fusion protein
(FAS/ER) mRNA, partial cds.
ACCESSION AF172068
VERSION AF172068.1 GI:5825505

KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1218)
AUTHORS Ye, Q., Chung, L.W., Li, S., and Zhou, H.Y.
TITLE Identification of a novel FAS/ER-alpha fusion transcript expressed in human cancer cells

JOURNAL Biochim. Biophys. Acta 1493 (3), 373-377 (2000)
MEDLINE 20472057
PUBMED 11018265
REFERENCE 2 (bases 1 to 1218)
AUTHORS Ye, Q.N., Li, S.W., Chung, L.W.K. and Zhou, H.Y.
TITLE Direct Submission
JOURNAL Submitted (25-JUL-1999) Department of Urology, University of Virginia, 1300 Jefferson Park Avenue, Charlottesville, VA 22903,
USA

FEATURES
source
Location/Qualifiers
1. .1218
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="393"
/cell_line="ARCap"
/cell_type="prostate cancer"
1. .1218
/gene="FAS/ER"
1. .1218

gene
CDS

/gene="FAS/ER"
/codon_start=1
/product="fatty acid synthase/estrogen receptor fusion protein"
/protein_id="AA053275.1"
/db_xref="GI:5825505"
/translation="SLARTSSAEYNNVNLVSPVLFQALWHVPEHAVVLEIAPHALLQ
AVLKRGLKPSCTIIPLMKKDRDNLEFFLAGIRRLHLSGIDANFNALFPVPEFPAPRG
TPLISPLIKWDSIALSLTADQMSVALLDAEPPILYSEYDTPRPFSEASMMGLLTNLA
DRELVHMINKAKRVPGVDLTLHDQVHLLCEAMLEILIMIGLVWRSMHPGKLLFAPNL
LLDRNQKCEVGMVEIFDMLLATSSRFMMNLQGEERFVCLKSIILLNSGVYTFLSSTL
KSLSEKHHRVLDKIDTLTLLHMAKAGLTQQOORLAQLLLILSHIRHMSNKGMEH
LYSMCKNVVPLDYDLLEMLDAHRLHAPTSRGASVEETDQSHLATAGTSSSHSLQKY
YITGEAEGRPATV"

ORIGIN

Query Match 100.0%; Score 41; DB 9; Length 1218;
Best Local Similarity 100.0%; Pred. No. 8.7e-07;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TTGTGGATGCTGAGCCCCCATACTCTATTCCGAGTATGA 41
|||||
Db 385 TTGTGGATGCTGAGCCCCCATACTCTATTCCGAGTATGA 425

RESULT 4

LOCUS AF172069 1218 bp mRNA linear PRI 06-SEP-1999
DEFINITION Homo sapiens fatty acid synthase/estrogen receptor fusion protein
(FAS/ER) mRNA, partial cds.
ACCESSION AF172069
VERSION AF172069.1 GI:5825508

KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1218)
AUTHORS Ye, Q.N., Li, S.W., Chung, L.W.K. and Zhou, H.Y.
TITLE Direct Submission
JOURNAL Submitted (25-JUL-1999) Department of Urology, University of Virginia, 1300 Jefferson Park Avenue, Charlottesville, VA 22908,
USA

FEATURES
source
Location/Qualifiers
1. .1218
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="73"
/cell_line="LNCap"
/cell_type="prostate cancer"
1. .1218
/gene="FAS/ER"
1. .1218
/gene="FAS/ER"
/codon_start=1
/product="fatty acid synthase/estrogen receptor fusion protein"
/protein_id="AA053276.1"
/db_xref="GI:5825509"

gene

CDS

ORIGIN

Query Match 100.0%; Score 41; DB 9; Length 1218;
Best Local Similarity 100.0%; Pred. No. 8.7e-07;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

ORIGIN

/db xref="taxon:32630"
/focus
697..1644
/organism="Homo sapiens"
/mol_type="mRNA"
/db xref="taxon:9606"
1..1644
/codon_start=1
/transl_table=11
/product="ECFP-estrogen receptor ligand binding domain
fusion protein"
/protein_id="AAF65451.1"
/db_xref="GI:7638249"
/translation="MWSKGELFTGVVPIVLVDGVNHFVSVSGEGDGYTKLT
LKFICTGKLPVPTLVITLWTGQCFSPDENKQDFKSPMPGYVQERTIFPK
DQNGYKRAEYFEGDTLVNRIELKIDFKEDGNILHGKYEYINSHVYITADKQKN
GIKAHFIRNIEDSGVADHYQONTPIGDGPVLLPDNHYLSQSALSKDPNKRDRH
MVLLEFVTAARMPGSGDMRAANLWPSLMIKRSKNSLALSLTADQMSALLDAEP
PILYSEVDPTRFSEASMMGLLTNLADRELVMINWAKRVFVDLTLDQVHLLLECA
WLEIMIGLWRSMEHPVKLLFAPNLLDNRQKCEVGMVEIFDMLLTSRFRMNL
QGEFVCLUKSIIILNSGVYFSLSTLSLEKDHHRVLDKITDTLIHLMAKAGITLQ
QQHQLAQLLILSHIRMSNKGMEHLYSMCKNVVPLYDILLLEMLDAHRLHAPTSGR
GASVEETDQSHLATAGTSSSHSLQKYITGEAGFPATV"

ORIGIN

Query Match 100.0%; Score 41; DB 12; Length 1644;
Best Local Similarity 100.0%; Pred. No. 8.7e-07;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TTGTTGGATGCTGAGCCCCCATACTCTATTCCGAGTATGA 41
DB 811 TTGTTGGATGCTGAGCCCCCATACTCTATTCCGAGTATGA 851

RESULT 9
AX824420
LOCUS AX824420 1770 bp DNA linear PAT 11-DEC-2003
DEFINITION Sequence 6 from Patent WO03070975.
ACCESSION AX824420
VERSION AX824420.1 GI:39750420
KEYWORDS synthetic construct
SOURCE synthetic construct
ORGANISM artificial sequences.
REFERENCE 1
AUTHORS Verdonk,G.N., Dijkema,R.N. and Schoonen,W.G.
TITLE An estrogen receptor alpha construct for use in a yeast 2-hybrid
JOURNAL assay
Patent: WO 03070975-A 6 28-AUG-2003;
Akzo Nobel N.V. (NL)
FEATURES
source
1..1770
Location/Qualifiers
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db xref="taxon:32630"
/note="Fusion DNA sequence page 17-19"

ORIGIN

Query Match 100.0%; Score 41; DB 6; Length 1770;
Best Local Similarity 100.0%; Pred. No. 8.8e-07;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TTGTTGGATGCTGAGCCCCCATACTCTATTCCGAGTATGA 41
DB 937 TTGTTGGATGCTGAGCCCCCATACTCTATTCCGAGTATGA 977

RESULT 10
AX411751
LOCUS AX411751 1788 bp DNA linear PAT 14-JUN-2002
DEFINITION Sequence 1 from Patent WO0228175.
ACCESSION AX411751
VERSION AX411751.1 GI:21444272

KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Chambon,P. and Metzger,D.
TITLE Transgenic mouse for targeted recombination mediated by modified
Cre-er
JOURNAL Patent: WO 0228175-A 1 11-APR-2002;
ASS POUR LE DEV DE LA RECH (FR)
FEATURES
source
1..1788
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db xref="taxon:9606"
1..1788
note="unnamed protein product"
/codon_start=1
/protein_id="CAD35258.1"
/db_xref="GI:21444273"
/translation="MTMTLTKASGMALLHQIQNELEPLNRPQLKIFLERPLGEVYL
DSSXPVNVYPEGAAVEFNAAAANAQVVGOTGLPYGPGSEAAAFSGNGLGCFPLNS
VSPFLMLLHPPQLSPFLQPHGQGVXYLENEPSGYTVREAGPAPFYRPNSDNRQOG
GRERLASTNDKSWAMESAKETRYCAVNDYASGYHYGVWSECGCAFFKESIQQHND
YMCPTAQCTTIDNRKSKQACRLKCYEYGMKGGIRKRRGRGMLKHKQRDDCEG
RGEVSGADMFAANLWPSLMIKRSKNSLALSLTADQMSALLDAEPPILYSEVDP
RPFSEASMMGLLTNLADRELVMINWAKRVFVDLTLDQVHLLLECAWLEIMIGLW
RSEVHEPGKLLFAPNLLDNRQKCEVGMVEIFDMLLTSRFRMNLQGEFVCLUKS
IILNSGVYFSLSTLSLEKDHHRVLDKITDTLIHLMAKAGITLQOQHRLAQLL
LILSHIRMSNKGMEHLYSMCKNVVPLYDILLLEMLDAHRLHAPTSGASVEETDQS
HLATAGTSSSHSLQKYITGEAGFPATV"

ORIGIN

Query Match 100.0%; Score 41; DB 6; Length 1788;
Best Local Similarity 100.0%; Pred. No. 8.8e-07;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TTGTTGGATGCTGAGCCCCCATACTCTATTCCGAGTATGA 41
DB 955 TTGTTGGATGCTGAGCCCCCATACTCTATTCCGAGTATGA 995

RESULT 11
AX824417
LOCUS AX824417 1803 bp DNA linear PAT 11-DEC-2003
DEFINITION Sequence 3 from Patent WO03070975.
ACCESSION AX824417
VERSION AX824417.1 GI:39750419
KEYWORDS synthetic construct
SOURCE synthetic construct
ORGANISM artificial sequences.
REFERENCE 1
AUTHORS Verdonk,G.N., Dijkema,R.N. and Schoonen,W.G.
TITLE An estrogen receptor alpha construct for use in a yeast 2-hybrid
JOURNAL assay
Patent: WO 03070975-A 3 28-AUG-2003;
Akzo Nobel N.V. (NL)
FEATURES
source
1..1803
Location/Qualifiers
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db xref="taxon:32630"
/note="Fusion DNA p 14-16"

ORIGIN

Query Match 100.0%; Score 41; DB 6; Length 1803;
Best Local Similarity 100.0%; Pred. No. 8.8e-07;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TTGTTGGATGCTGAGCCCCCATACTCTATTCCGAGTATGA 41

```

Db      970 TTGTTGGATGCTGAGCCCCCATCTCTATTCCGAGTATGA 1010
RESULT 12
AR116270
LOCUS      AR116270      1956 bp      DNA      linear      PAT 16-MAY-2001
DEFINITION Sequence 2 from patent US 6133027.
ACCESSION AR116270
VERSION    AR116270.1 GI:14096592
KEYWORDS
SOURCE     Unknown.
ORGANISM   Unclassified.
REFERENCE  1 (bases 1 to 1956)
AUTHORS    Yee, J.-K., Friedmann, T. and Chen, S.-T.
TITLE      Inducible expression system
JOURNAL    Patent: US 6133027-A 8/17/OCT-2000
FEATURES   source
            1..1956
            /organism="unknown"
            /mol_type="unassigned DNA"

ORIGIN
Query Match      100.0%; Score 41; DB 6; Length 1956;
Best Local Similarity 100.0%; Pred. No. 8.8e-07;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 TTGTTGGATGCTGAGCCCCCATCTCTATTCCGAGTATGA 41
        |||||
        1123 TTGTTGGATGCTGAGCCCCCATCTCTATTCCGAGTATGA 1163

Db      1123 TTGTTGGATGCTGAGCCCCCATCTCTATTCCGAGTATGA 1163

RESULT 13
AX411753
LOCUS      AX411753      1983 bp      DNA      linear      PAT 14-JUN-2002
DEFINITION Sequence 3 from Patent WO0228175.
ACCESSION AX411753
VERSION    AX411753.1 GI:21444274
KEYWORDS
SOURCE     synthetic construct
            Location/Qualifiers
            1..1983
            /organism="synthetic construct"
            /mol_type="unassigned DNA"
            /db_xref="taxon:32630"

ORGANISM   synthetic construct
            artificial sequences.

REFERENCE  1
AUTHORS    Chambon, P. and Metzger, D.
TITLE      Transgenic mouse for targeted recombination mediated by modified
            Cre-er
JOURNAL    Patent: WO 0228175-A 3 11-APR-2002;
            ASS FOUR LE DEV DE LA RECH (FR)
FEATURES   source
            Location/Qualifiers
            1..1983
            /organism="synthetic construct"
            /mol_type="unassigned DNA"
            /db_xref="taxon:32630"

CDS
            1..1983
            /note="unnamed protein product; Chimeric sequence
            Homosapiens-Bacteriophage P1"
            /codon_start=1
            /transl_table=11
            /protein_id="CAD35259.1"
            /db_xref="GI:21444275"
            /translation="MSNLLIVHQNLPALPVDATSDVRKNLMDMFRDQAFSEHTWKM
            LLSVCRSWAACKNNKWPFPEDVDYLLYLQARGLAVKTIQHLGQLNMLHRRS
            GLPSPDSNAVSLVMRIRKENVDAGERAKQALAFERTDFDQVRSIMENSRCODIRN
            LAFGLAYNTLLRIAEIARIRVKDIISRTDGRMLIHGRTKTLVSTAGVEKALSLGVT
            KLVERWISVGVADDPNNVLCVRKNGVAAPSATSQSLSTRALEGIFEATHRLIYGAK
            DDSGQYLAWSGHSARVGAARDWARAGVSIPEIMQAGGWTNNVIMNYIRNLDSGTGA
            MWLLDGDLEPSAGDMRAANLWPSPLMKRSKNSLALSLTADQVMSALLDAEPPIIL
            YSEYDTRPFSSEAMMGLLTNLADRELIVHMINWAKRVPGVFDTLHDQVHLLCAWLE
            ILMIGLWRSMEHPGKLLFAPNLLDRNOCKVEGVMEIFDMLLATSSRFPMNLOGE
            EFVCLAKSIILLNSGVVTFLLSSTLSLEEKDHIHRVLDKITDTLIHLMKAGLTQQOH
            ORLAQILLILSHIRHMSNKRMEHLYSMKCKNVVPLDYDLLLEADAHLHAPTSGGAS
            VEETDQSHLATAGTSSSHSLQKYIITGEAGFPATV"

ORIGIN
Query Match      100.0%; Score 41; DB 6; Length 1983;
Best Local Similarity 100.0%; Pred. No. 8.8e-07;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 TTGTTGGATGCTGAGCCCCCATCTCTATTCCGAGTATGA 41
        |||||
        1150 TTGTTGGATGCTGAGCCCCCATCTCTATTCCGAGTATGA 1190

Db      1150 TTGTTGGATGCTGAGCCCCCATCTCTATTCCGAGTATGA 1190

RESULT 15
AX411757
LOCUS      AX411757      1983 bp      DNA      linear      PAT 14-JUN-2002
DEFINITION Sequence 7 from Patent WO0228175.
ACCESSION AX411757
VERSION    AX411757.1 GI:21444278
KEYWORDS
SOURCE     synthetic construct
            Location/Qualifiers
            1..1983
            /organism="synthetic construct"
            /mol_type="unassigned DNA"
            /db_xref="taxon:32630"

ORGANISM   synthetic construct
            artificial sequences.

REFERENCE  1
AUTHORS    Chambon, P. and Metzger, D.

```

TITLE Transgenic mouse for targeted recombination mediated by modified
Cre-er
JOURNAL Patent: WO 0228175-A 7 11-APR-2002;
ASS POUR LE DEV DE LA RECH (FR)
FEATURES Location/Qualifiers
source
1..1983
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
1..1983
/note="unnamed protein product; Chimeric sequence
Homo sapiens-Bacteriophage P1"
/codon_start=1
/transl_table=11
/protein_id="CAD35261.1"
/db_xref="GI:21444279"
/translation="MSNLLTVHQNLPAFPVDTATSEVERKNLMDMFRDQAFSEHTWKM
LLSVCRSWAAWCKLNRRKWFPAEPEDVDYLLYLQAGLAVKTIQQHGLQNLMLHRRS
GLPRPSDNAYSLVNRIRKENVDAGERAKALAFERTDFQVRSLSMENSRCQDIRN
LAFGLIAYNTLLRIAEIARIKVDISRTDGGMLIHIGRTKTLVSTAGVEKALSLGVT
KLVERWISVSGVADDPNNYLCFRKNGVAPSAQSOLSTRALEGIFEATHRLIYGAK
DDSGQRYLAWSGHSARVGAARDMARAGVSIPEIMQAGGWTNNVNIYIRNLDSGTGA
NVRLEDDGLFESAGDMAANLWPSPLMIKRSKNSLALSUTADQMVSAALLDAEPPIL
YSEYDTPFPFSEASMMGLITNLADRELVHMINWAKRVFQFVDLTLDHQVHLECAWLE
ILMIGLVKRSMEHPGKLLFAPNLLDNRQKCGEVEIFDMLLATSSFRFMNLQGE
BEVCLKSIILLNSGYVTFSSTLKSLSEKDHTRVLDKITDTLIHLMAKAGLTQQQH
QRLAQILLILSHIRHMSKMEHLYSMCKKNVPLVDLLLEAADAHLHAPTSGGAS
VEETDQSHLATAGTSSSHSLQKYYITGAEGFPATA"

ORIGIN

Query Match 100.0%; Score 41; DB 6; Length 1983;
Best Local Similarity 100.0%; Pred No. 8.8e-07;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 TTGTTGGATGCTGAGCCCCCATCTCTATTCGAGTATGA 41
Db 1150 TTGTTGGATGCTGAGCCCCCATCTCTATTCGAGTATGA 1190

Search completed: November 3, 2004, 07:43:07
Job time : 257.809 secs

Search

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 3, 2004, 04:55:34 ; Search time 255.809 Seconds
(without alignments)
7579.407 Million cell updates/sec

Title: US-09-933-267a-1_COPY_64311_64351

Perfect score: 41

Sequence: 1 tctctcgtgactttgtccgt.....aattgataagatctaatttg 41

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 4526729 seqs, 23644849745 residues

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*
1: gb_ba.*
2: gb_ptg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	41	100.0	56124	9	AB090237 Homo sapi
2	41	100.0	180832	2	AC044828 Homo sapi
3	41	100.0	349980	6	AX232503 Sequence
4	41	100.0	349980	6	AX453703 Sequence
5	39.4	96.1	110000	6	AR397408 Sequence
6	39.4	96.1	119737	9	AL035695 Human DNA
7	26.6	64.9	210954	5	AJ627213 Gallus ga
8	25	61.0	179906	2	AC119339 Rattus no
9	24	58.5	614	14	AF065554 HIV-1 iso
10	24	58.5	131823	2	AC010497 Homo sapi
11	24	58.5	134349	9	AC010296 Homo sapi
12	24	58.5	154061	9	AC008524 Homo sapi
13	24	58.5	179647	2	AC108124 Homo sapi
14	24	58.5	188207	2	AC010233 Homo sapi
15	23.8	58.0	1347	6	AR394374 Sequence
16	23.8	58.0	86761	10	AL663044 Mouse DNA
17	23.8	58.0	213038	2	AC11823 Rattus no
18	23.8	58.0	235502	2	AC103222 Rattus no
19	23.8	58.0	300225	1	AE016949 Enterococ

C 20	23.6	57.6	151341	10	AC115236
C 21	23.6	57.6	181157	10	AC121789
C 22	23.6	57.6	207528	2	AC139331
C 23	23.6	57.6	245877	2	AC128382
C 24	23.6	57.6	260121	2	AC130738
C 25	23.6	57.6	270188	2	AC112821
C 26	23.6	57.6	338614	2	AC103268
C 27	23.4	57.1	591	6	CQ103678
C 28	23.4	57.1	591	6	CQ142508
C 29	23.4	57.1	591	6	CQ225743
C 30	23.4	57.1	591	6	CQ263783
C 31	23.4	57.1	591	6	CQ300858
C 32	23.4	57.1	10585	1	AE006124
C 33	23.4	57.1	110000	2	BX294176
C 34	23.4	57.1	150831	9	AC007719
C 35	23.4	57.1	171263	10	AC118702
C 36	23.4	57.1	184674	2	AC102557
C 37	23.4	57.1	192449	10	AC115884
C 38	23.4	57.1	207870	10	AL669900
C 39	23.4	57.1	235317	2	AC121371
C 40	23.4	57.1	238007	2	AC107569
C 41	23.4	57.1	241571	2	AC108281
C 42	23.2	56.6	123947	9	AC005045
C 43	23.2	56.6	168745	2	AC119552
C 44	23.2	56.6	248329	2	AC109989
C 45	23	56.1	609	11	BV079087

ALIGNMENTS

RESULT 1
AB090237
LOCUS
DEFINITION Homo sapiens gene for estrogen receptor alpha, exon 1a, 1b, 2.
ACCESSION
AB090237
VERSION
AB090237.1
KEYWORDS
SOURCE
Homo sapiens (human)

Seq ID #1 fr

REFERENCE
AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

mRNA

exon

mRNA

exon

AB090237 56124 bp DNA linear PRI 12-APR-2003
Homo sapiens gene for estrogen receptor alpha, exon 1a, 1b, 2.

AB090237 Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

Inoue, A., Hayashi, S., Aoyagi, K., Nishigaki, M., Sasaki, H. and
Kiyama, R.
A reporter gene assay for evaluation of tissue-specific responses
to estrogens based on the differential use of promoters A to F of
the human estrogen receptor alpha gene

Unpublished
22516287
12628303
2 (bases 1 to 56124)
Submitted 20-AUG-2002
Direct Submission
Sasaki, H.

Research Institute, Genetics Division, 1-1, Tsukiji 5-chome,
Chuo-ku, Tokyo 104-0045, Japan (E-mail: hksasaki@gen2.res.ncc.go.jp,
Tel: 81-3-3542-2511 (ex. 4402), Fax: 81-3-3541-2685)
Location/Qualifiers
1. 56124
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
join(7128..7219,53065..53195)
/product="estrogen receptor alpha"
/note="generated by promoter E and alternative splicing"
7128..7219
/product="estrogen receptor alpha"
/number=1a
join(41744..41866,53065..53195)
/product="estrogen receptor alpha"
/note="generated by promoter F and alternative splicing"
41744..41866

/product="estrogen receptor alpha"
/number=1b
53065..53195
/product="estrogen receptor alpha"
/number=2

ORIGIN

Query Match 100.0%; Score 41; DB 9; Length 56124;
Best Local Similarity 100.0%; Pred. No. 8.1e-05; Indels 0; Gaps 0;
Matches 41; Conservative 0; Mismatches 0;
QY 1 TCTCTCCTGACTTTCGCGTAATTCGTAAGATCTAAATTTG 41
Db 53226 TCTCTCCTGACTTTCGCGTAATTCGTAAGATCTAAATTTG 53266

RESULT 2

AC044828/c
LOCUS AC044828 180832 bp DNA linear HTG 20-SEP-2000
DEFINITION Homo sapiens chromosome 6 clone RP11-282P11 map 6, WORKING-DRAFT
SEQUENCE, 34 unordered pieces.
AC044828
ACCESSION AC044828.2 GI:10198537
VERSION HTG; HTGS PHASE1; HTGS_DRAFT.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 180832)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 6, clone RP11-282P11

REFERENCE

1 (bases 1 to 180832)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavskiy,L., Boukhalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Collangelo,M., Collins,S.,
Collamore,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lechoczy,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheters,R.,
Meldrum,J., Mereus,L., Minova,T., Miranda,C., Mienga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.

TITLE

Submitted (12-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Sep 20, 2000 this sequence version replaced gi:7543791.
All repeats were identified using RepeatMasker:
Smith, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

JOURNAL

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L8038
Center clone name: 282 P.11
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731

Consensus quality: 162722 bases at least Q40
Consensus quality: 171198 bases at least Q30
Consensus quality: 174912 bases at least Q20
Insert size: 185000; agarose-fp
Insert size: 177532; sum-of-contigs
Quality coverage: 4.1 in Q20 bases; agarose-fp
Quality coverage: 4.3 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 34 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1
* 3320: contig of 3320 bp in length
* 3420: gap of 100 bp
* 3421: 4535: contig of 1115 bp in length
* 4536: gap of 100 bp
* 4537: 5881: contig of 1246 bp in length
* 5882: gap of 100 bp
* 5883: 7889: contig of 1908 bp in length
* 7890: gap of 100 bp
* 7891: 9338: contig of 1349 bp in length
* 9339: gap of 100 bp
* 9340: 10913: contig of 1475 bp in length
* 10914: gap of 100 bp
* 11014: 13206: contig of 2193 bp in length
* 13207: gap of 100 bp
* 13208: 14999: contig of 1693 bp in length
* 15000: gap of 100 bp
* 15001: 18000: contig of 2901 bp in length
* 18002: gap of 100 bp
* 18003: 20868: contig of 2768 bp in length
* 20869: gap of 100 bp
* 20870: 24646: contig of 3678 bp in length
* 24647: gap of 100 bp
* 24648: 28762: contig of 4016 bp in length
* 28763: gap of 100 bp
* 28764: 31776: contig of 2914 bp in length
* 31777: gap of 100 bp
* 31778: 35418: contig of 3542 bp in length
* 35419: gap of 100 bp
* 35420: 40331: contig of 4713 bp in length
* 40332: gap of 100 bp
* 40333: 44448: contig of 4117 bp in length
* 44449: gap of 100 bp
* 44450: 49300: contig of 4752 bp in length
* 49301: gap of 100 bp
* 49302: 53231: contig of 3831 bp in length
* 53232: gap of 100 bp
* 53233: 57997: contig of 4366 bp in length
* 57998: gap of 100 bp
* 57999: 62107: contig of 4310 bp in length
* 62108: gap of 100 bp
* 62109: 65070: contig of 2863 bp in length
* 65071: gap of 100 bp
* 65072: 70534: contig of 5364 bp in length
* 70535: gap of 100 bp
* 70536: 77999: contig of 7365 bp in length
* 77999: gap of 100 bp
* 78000: 83751: contig of 5651 bp in length
* 83752: gap of 100 bp
* 83753: 90537: contig of 6687 bp in length
* 90538: gap of 100 bp
* 90539: 97483: contig of 6746 bp in length
* 97484: gap of 100 bp
* 97485: 106001: contig of 8518 bp in length
* 106002: gap of 100 bp
* 106003: 116583: contig of 10482 bp in length
* 116584: gap of 100 bp
* 125776: contig of 9093 bp in length

* 125777 125876: gap of 100 bp
* 125877 131540: contig of 5664 bp in length
* 131541 131640: gap of 100 bp
* 131641 141557: contig of 9817 bp in length
* 141558 141657: gap of 100 bp
* 141658 154976: contig of 13319 bp in length
* 154977 155076: gap of 100 bp
* 155077 167892: contig of 12816 bp in length
* 167893 167992: gap of 100 bp
* 167993 180832: contig of 12840 bp in length.

FEATURES

Location/Qualifiers
1..180832
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="6"
/map="6"
/clone="RP11-282P11"
/clone_lib="RPC1-11 Human Male BAC"
1..3320
/note="assembly_fragment"
clone_end:SP6
vector_side:left
3421..4535
/note="assembly_fragment"
4636..5881
/note="assembly_fragment"
5982..7889
/note="assembly_fragment"
7990..9338
/note="assembly_fragment"
9439..10913
/note="assembly_fragment"
11014..13206
/note="assembly_fragment"
13307..14999
/note="assembly_fragment"
15100..18000
/note="assembly_fragment"
18101..20868
/note="assembly_fragment"
20969..24646
/note="assembly_fragment"
24747..28762
/note="assembly_fragment"
28863..31776
/note="assembly_fragment"
31877..35418
/note="assembly_fragment"
35519..40231
/note="assembly_fragment"
40332..44448
/note="assembly_fragment"
44549..49300
/note="assembly_fragment"
49401..53231
/note="assembly_fragment"
53332..57697
/note="assembly_fragment"
57798..62107
/note="assembly_fragment"
62208..65070
/note="assembly_fragment"
65171..70534
/note="assembly_fragment"
70635..77999
/note="assembly_fragment"
78100..83750
/note="assembly_fragment"
83851..90537
/note="assembly_fragment"
90638..97383

Query Match 100.0%; Score 41; DB 2; Length 180832;
Best Local Similarity 100.0%; Pred. No. 7.8e-05;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCTCTCGTACTTTGTCGTAATAATTGATAAGATCTAAATTG 41
|||||
Db 108986 TCTCTCGTACTTTGTCGTAATAATTGATAAGATCTAAATTG 108946
|||||

RESULT 3

AX232503
LOCUS AX232503 349980 bp DNA linear PAT 11-SEP-2001
DEFINITION Sequence 1 from Patent WO0162969.
ACCESSION AX232503
VERSION AX232503.1 GI:15592567
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Kalush, F., Cassel, M.J., Hwang, S.S. and Winn-Deen, E.S.
TITLE Estrogen receptor alpha variants and methods of detection thereof
JOURNAL Patent: WO 0162969-A 1 30-AUG-2001;
PE Corporation (NY) (US)
FEATURES
Location/Qualifiers
source 1..349980
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
/note="Original sequence 1: Length 465237 has been
replaced by-seq 1: from 1 to 349980 349980 bases-seq 3:
from 300001 to 465237 165237 bases"

ORIGIN

Query Match 100.0%; Score 41; DB 6; Length 349980;
Best Local Similarity 100.0%; Pred. No. 7.6e-05;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCTCTCGTACTTTGTCGTAATAATTGATAAGATCTAAATTG 41
|||||
Db 64311 TCTCTCGTACTTTGTCGTAATAATTGATAAGATCTAAATTG 64351
|||||

RESULT 4

AX453703
LOCUS AX453703 349980 bp DNA linear PAT 06-JUL-2002
DEFINITION Sequence 1 from Patent WO0234945.
ACCESSION AX453703
VERSION AX453703.1 GI:21712921
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Kalush, F., Cassel, M.J., Hwang, S.S. and Winn-Deen, E.S.
TITLE Estrogen receptor alpha variants and methods of detection thereof
JOURNAL Patent: WO 0234945-A 1 02-MAY-2002;
Applera Corporation (US)
FEATURES
Location/Qualifiers
source 1..349980
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
/note="original length seq 1: 465237-replaced by seq 1:
000.001 -> 349.980-seq 4: 300.001 -> 465237"

ORIGIN

Query Match 100.0%; Score 41; DB 6; Length 349980;
Best Local Similarity 100.0%; Pred. No. 7.6e-05;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCTCTCCGACTTTCGCGTAATTAAGATCTAAATTTG 41
 |||||
 Db 64311 TCTCTCCGACTTTCGCGTAATTAAGATCTAAATTTG 64351

RESULT 5
 AR397408_0
 WFCOMMENT
 Sequence split into 4 fragments LOCUS AR397408 Accession AR397408
 Fragment Name Begin End
 AR397408_0 1 110000
 AR397408_1 100001 210000
 AR397408_2 200001 310000
 AR397408_3 300001 392000
 LOCUS AR397408 PAT 18-DEC-2003
 DEFINITION Sequence 11 from patent US 6617162.
 ACCESSION AR397408
 VERSION AR397408.1 GI:40134206
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 UNCLASSIFIED.
 REFERENCE 1 (bases 1 to 392000)
 AUTHORS Dobie, K.W. and Roach, M.P.
 TITLE Antisense modulation of estrogen receptor alpha expression
 JOURNAL Patent: US 6617162-A 11 09-SEP-2003;
 LOCATION/Qualifiers
 1. .392000
 /organism="unknown"
 /mol_type="genomic DNA"

ORIGIN
 Query Match 96.1%; Score 39.4; DB 6; Length 110000;
 Best Local Similarity 97.6%; Pred. No. 0.00031;
 Matches 40; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCTCTCTGACTTTGCGGTAATTAAGATCTAAATTTG 41
 |||||
 Db 12773 TCTCTCTGACTTTGCGGTAATTAAGATCTAAATTTG 12813

RESULT 6
 HS404G5/c
 LOCUS
 DEFINITION Human DNA sequence from clone RP3-404G5 on chromosome 6q24.1-25.2, complete sequence.
 ACCESSION AL035695
 VERSION AL035695.17 GI:5668650
 KEYWORDS HTG.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 119737)
 Direct Submission
 Submitted (04-MAR-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
 On Jul 31, 1999 this sequence version replaced gi:5650586.
 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
 The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:
 Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at
 http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping

Group. Further information can be found at
 http://www.sanger.ac.uk/HGP/Chr6
 ----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: http://www.sanger.ac.uk
 Contact: humquerry@sanger.ac.uk

 This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.
 RP3-404G5 is from the library RP3-3 constructed by the group of Pieter de Jong. For further details see
 http://www.chori.org/bacpac/home.htm
 VECTOR: pCYPAC2.

 Location/Qualifiers
 1. .119737
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="6"
 /map="q24.1-25.2"
 /clone="RP3-404G5"
 /clone_lib="RP3-3"

FEATURES
 Source
 1. .119737
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="6"
 /map="q24.1-25.2"
 /clone="RP3-404G5"
 /clone_lib="RP3-3"

ORIGIN
 Query Match 96.1%; Score 39.4; DB 9; Length 119737;
 Best Local Similarity 97.6%; Pred. No. 0.00031;
 Matches 40; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCTCTCTGACTTTGCGGTAATTAAGATCTAAATTTG 41
 |||||
 Db 56366 TCTCTCTGACTTTGCGGTAATTAAGATCTAAATTTG 56326

RESULT 7
 AJ627213/c
 LOCUS
 DEFINITION Gallus gallus primary C-C chemokine receptor cluster.
 ACCESSION AJ627213
 VERSION AJ627213.1 GI:42557656
 KEYWORDS C-C chemokine receptor 11 like; C-C chemokine receptor 5 like; C-C chemokine receptor 8 like; C-C chemokine receptor 9 like; CCRI-L gene; CCR5-L gene; CCR8-L gene; CCR9-L gene; CCXC chemokine receptor 1 like; CCXCR1-L gene; FCY01 gene; FYVE and coiled-coil.
 SOURCE Gallus gallus (chicken)
 ORGANISM Gallus gallus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
 1
 Morroll, S. and Bumstead, N.
 Identification and characterisation of the primary C-C chemokine receptor gene cluster in chicken
 Unpublished
 2 (bases 1 to 210954)
 Morroll, S.
 Direct Submission
 Submitted (10-FEB-2004) Morroll S., Avian Genetics, Institute for Animal Health, Compton Lab., Compton, Newbury, Berks, RG20 7NN, UNITED KINGDOM
 Location/Qualifiers
 1. .210954
 /organism="Gallus gallus"
 /mol_type="genomic DNA"
 /db_xref="taxon:9031"
 /chromosome="2"
 complement(24094..24960)

FEATURES
 source
 1. .210954
 /organism="Gallus gallus"
 /mol_type="genomic DNA"
 /db_xref="taxon:9031"
 /chromosome="2"
 complement(24094..24960)

gene

```

CDS
/gene="CCR1-L"
/complement (24094..24960)
/gene="CCR1-L"
/codon_start=1
/evidence=experimental
/product="C-C chemokine receptor 11 like"
/protein_id="CAF28776.1"
/db_xref="GI:42557657"
/translation="MTDIYLLNLAISDLLIFSLPFWAYAAHWDIFGDALCRILSGV
YLLGFSYGIFFIILTVDRYLAIVHAFALKARTVTVGILTSVMAVAFASVPGIV
PHKQEHRTVTCSAHYQBOEDWKQPLAKMNLGLVIMIMICSYQILIKTLQ
CRNEKNKAVRLIFIMVYFFPWAPYINICILLRDFQDSFITSCEISGQKQATQVT
ETISMIHCCINPVITAFAGEXPEKRYLRSFRFRQIASHPFSKYCPVFAADTVERASTYT
QSTGEQVSAAL"
/complement (41032..42096)
/gene="CCR5-L"
/complement (41032..42096)
/gene="CCR5-L"
/codon_start=1
/product="C-C chemokine receptor 5 like"
/protein_id="CAF28777.1"
/db_xref="GI:42557658"
/translation="WENYTDLDGDMPTTTFDYGDTPCMGTKEKHPAANFLPPLVSLV
IPFGINLLVLLVYKYLKMSMDIYLNLAISDLLFVSLPFWAYAAHWDIFGDA
LCRLSGVILGFSYGIFFIILTVDRYLAIVHAFALKARTVTVGILTSVMAVAF
FASVPGIVFHKTQENTQCTCSFHPSPDALINWQHSYILKNLGLIIMIPCSY
QILRTLFCGRNEKKQAVRLIFVIMIFYFIPTWPHVASFVHTFOTSPFSDCDSQR
LEKTIQVETTSVMHCINPVYAFVGEKFRKYLHMPFRKVAHDLCKKCSLYREKL
ERVSTFTASTAEHDISTGL"
/complement (60711..61844)
/gene="CCR8-L"
/complement (60711..61844)
/gene="CCR8-L"
/codon_start=1
/product="C-C chemokine receptor 8 like"
/protein_id="CAF28778.1"
/db_xref="GI:42557659"
/translation="MEQRKPTGRHTRALYLVFFPFSQBSKMNPTDLFLSTYDGYD
ENTAPCNEGNSFFPRKSLFLLYLVFVCLLNGSLVLLTLLTKRLMTWTDICLN
LAASDLLFVLPQAYVASDQVFGNALCKIMAGIYVTVGYSSIIFRITMSIDRYIA
IVHAYAMKIRTSACGMVLSILVNLVAGLASVNLVFNQLEIEQSQCVCVPIYPPGN
IKWYTPAANILGILLPFLSLIHCYALQILVNLNRKKNQNKIKAMFIFIIVIFLFF
WTPNVVLFDLSQLSLIIDNCQASSQITLALQLETISFIHCCINPVYAFAGVTEK
AHLKRLQPCARILWSTRGSGVTQSSLVLSIQSGSDSAGVL"
/gene="CCR9-L"
/gene="CCR9-L"
/codon_start=1
/product="CCXC chemokine receptor 1 like"
/protein_id="CAF28779.1"
/db_xref="GI:42557660"
/translation="WDREQYPSGDDNVSPEYVLNSENVCVMGNFYFIYFTTTLVLT
LAELLSLLNTLVMLFKENLSLNFIMNLCISDLVFCMLPFWADQTFGWIF
GEFLCKAVNAFTSGYGVFFLTLMLIRLSVSVSLTSQTCYCSGLVCLLWTF
CSLIVVPEMTHTVTLTEVSTCYDDMKKVDIYQRNILFLISFGIIFCYINI
LIILRLRSRKRKTKVLIIVVAVFSLWAPYNLFLITFPPTCQYKEDTLTLLAFH
ISRKIAFSHCCNLPVLVAVGKPKSHLFRICGQYLPCCCDEVSIGSQAKFHEDAEI
Y"
/gene="CCR9-L"
/gene="CCR9-L"
/codon_start=1
/product="FYVE and coiled-coil"
/protein_id="CAF28780.1"
/db_xref="GI:42557661"
/translation="MRVELQALRQKELQSIHOLEMNEQLQAAVLSQLEQLRPEK
EKNNGVSEENSLPKMITELQKCEVSHSTQVHDLQKQSLQSLNVAQVQKEYSYK
LAQATSKEDYAKLQILNLEFVSRLVAMKELIDELVAKLSSTQKLNLIARVD
AALEBKHQMAQCDSALQIQALLERKLOQTEKEAENQRUSDECTSLQTAEEQLRLK
EEAQKELIESNYCLNLTARREGSKLRSLETMEKEVDALQKALTLEKKNMAELQTQVM
ESLAQVGSLEKDLFEARKEKEKLEEYGKMEELKEPAQSAQKFGQEGHLLKVSFT
VCSLEQKRLLYEKEHLSQKVELEQMRQONSTVNEMSESRKLTENVDLQOSKK
KVEEKLNLKASQSLAEAVARLRASEQLQSEIIDALVSVDEKELKRSQNKQJDED
IQNARRQSLLEKLEALQSDYREKEREETTRSYASLEGQLKSAKQHSIQVKSILN
TKKESKESLQSLAEKEIQLGMECCQQLRKEARHRRKAEETLEKLSAENTCLQQL
TKLJESLTXSEKSEKHQQAASLEKADKELASRLTVSEQLQNRDEVSRLQTEVL
DLRYLQOTDEREQLSLAITETVLGEQVLVQOLKEOTSLNHNHVOELVQCKER
EVLKREQEAVVLQKTELENNLLSLKEELSKFQYLEAARMENVENKLLHPTNTDMA
ELGQICALSSEKPYDAEQLQAKERLEKEQAAQOEKLDKHDLSNRQENRSQLEK
LEBAQICVSAVPSLQALQETAKKQSQFETSQEBLSAIFQNSTEILNYQTKFKVN"
/complement (136644..137792)
/gene="CCR9-L"
/complement (136644..137792)
/gene="CCR9-L"
/codon_start=1
/product="C-C chemokine receptor 9 like"
/protein_id="CAF28781.1"
/db_xref="GI:42557662"
/translation="MTSLDYRNDASGLSVIGNPINDTLMCDRRQVQWQAFAPLPVF
FWLIPFVGTGVALVLYKYRFRSMDRYLHLVADLILLFTLPFWATASSGM
IFRNFMCVNVSMVKINPYGCLIFCTCSFPRYLTIQVATKAKSKQRRIIRSKVVCF
AVWLASVLCPLPEIMYSQSKQIGAVTCVMYTPENIAGFRAVLVLDVTKVGFPLL
VWCYITLIIHLQAKRCQKHSKLIITMIITAPLQSOPFYNTVILIKITINMYTGAV
YSCQTINGLDIGLQVTSIAFLHSLCLNPLFYFAGERPRMALRMVOSTGRYLLGGQD
QCSLIGDSQESHSSNWFAMLGRRVRNLSLTSLTNLASSWVPAQCQV"
ORIGIN
Query Match 64.9%; Score 26.6; DB 5; Length 210954;
Best Local Similarity 78.0%; Pred. No. 16;
Matches 32; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
QY 1 TCTCTCTGATCTTTCCTCGTAAATGATAAGATCTAAATTTG 41
| | | | | | | | | | | | | | | | | | | | | |
Db 109834 TTTCTTAATCTTACCTTAAATTTGAAAGATCTAAATTTG 109794
| | | | | | | | | | | | | | | | | | | | | |

RESULT 8
AC119339
LOCUS Rattus norvegicus clone CH230-334F22, WORKING DRAFT SEQUENCE.
DEFINITION AC119339
ACCESSION AC119339
VERSION AC119339.5 GI:25138065
HTG; HTGS PHASE2; HTGS DRAFT; HTGS FULLTOP.
KEYWORDS Rattus norvegicus (Norway rat)
SOURCE Rattus norvegicus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 179906)
Muzny, D., Marie, Metzker, M., Lee, Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
Aryalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, B.,
Cardenas, J., Carter, K., Cavazos, I., Cesar, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, X., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Deigado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Gebregregis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hoques, M.,
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorensueta, L., Loulseg, H., Lozano, R.J., Lu, X., Ma, J.,

```

Maheshwari, M., Mahindartine, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawlin, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Ngunidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokelimeh, O., Okwuonu, G., Olarnpunaagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfankoch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajls, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, P., Williams, G., Willison, R., Wleczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.
 Direct Submission
 Unpublished
 2 (bases 1 to 179906)
 Worley, K.C.
 Direct Submission
 Submitted (26-APR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 179906)
 Rat Genome Sequencing Consortium.
 Direct Submission
 Submitted (20-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

On Nov 20, 2002 this sequence version replaced gi:23616725.
 The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GUXF

Center clone name: CH230-334F22

----- Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 165431 bases at least Q40

Consensus quality: 166914 bases at least Q30

Consensus quality: 168010 bases at least Q20

Estimated insert size: 170562; sum-of-contigs estimation

Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 1 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes

* of the gaps between them are based on estimates that have
 * provided by the submitter.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 * 1 179906: contig of 179906 bp in length.

FEATURES

source

1..179906

/organism="Rattus norvegicus"

/mol_type="genomic DNA"

/db_xref="taxon:10116"

/clone="CH230-334F22"

1..1691

/note="wgs end extension"

clone_end:Sp6"

complement(2760..3618)

/note="clone_boundary"

clone_end:Sp6"

site:

end sequence:B2258948"

159271..161083

/note="wgs contig"

misc_feature

173735..174575

/note="clone_boundary"

clone_end:T7

site:

end sequence:B2258947"

178556..179906

/note="wgs end extension"

misc_feature

clone_end:T7"

ORIGIN

Query Match 61.0%; Score 25; DB 2; Length 179906;
 Best Local Similarity 75.6%; Pred. No. 64;
 Matches 31; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1 TCTCTCTGACTTTGCCGTAATTCATAGATCTTAATTG 41

Db 96779 TGCTTCTGCTCTGCTGTTTAATGATAGAAATTAATTG 96819

RESULT 9

AF065554/c

LOCUS

DEFINITION AF065554 614 bp DNA linear VRL 11-JUN-1998
 HIV-1 isolate 19-108V2 from the USA, envelope glycoprotein C2-V5
 region (env) gene, partial cds.

ACCESSION

AF065554

VERSION

AF065554.1 GI:3211886

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

1..614

/organism="Human immunodeficiency virus 1"

```

/mol_type="genomic DNA"
/isolate="19-108V2"
/db_xref="taxon:11676"
/country="USA"
/notes="sample from mother; this mother had an infant who
initially had a positive PCR test for the HIV virus,
subsequently the infant seroreverted and had repeated
negative PCR tests; sequences from the infant's sporadic
positive PCR test (see GenBank Accession Numbers AF065555
and AF065556) and from the mother were compared and found
to be unrelated; Frenkel et al's interpretation of the
infant's initial, positive PCR test is that it was caused
by contaminants or other laboratory errors, and the infant
had not in fact been infected; other researchers have
interpreted this as reflective of a transient infection"
<1..>614
/genes="env"
<1..>614
/genes="env"
/notes="C2-V5 region; gp120"
/codon_start=2
/product="envelope glycoprotein"
/protein_id="AAC21517.1"
/db_xref="GI:3211887"
/translation="EXEVIKRSNFTDNVKTIIIVLKKSVKINTRPNNTRKSIPIG
PGRAFYTGDIIGIRQAHNCISKVWNNTLKIIVRKLRFQVKNKTIIFNQSSGDP
IVLHFNCGGEFFVCNTELFNSTWNGNDTQINNTIGNDTFLPCRILKQIINWQEV
GKAWYAPPIRXKIRCSSNITGLLLTRDGNENGTEITRPGGG"

```

ORIGIN

```

Query Match      58.5%; Score 24; DB 14; Length 614;
Best Local Similarity 84.4%; Pred. No. 1.8e+02;
Matches 27; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

```

```

Qy 8 TGACTTTGTCGTAATAATGATAAGATCTAATT 39
    |||||
Db 47 TGACATTGTCGTAATAATTTTCAGATCTAATT 16
    |||||

```

```

RESULT 10
AC010497      131823 bp DNA linear HTG 04-OCT-2001
LOCUS Homo sapiens chromosome 5 clone CTD-2353F3, WORKING DRAFT SEQUENCE,
DEFINITION 3 ordered pieces.
ACCESSION AC010497
VERSION HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ACTIVEFIN.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 131823)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 5
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 131823)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (15-SEP-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Jul 1, 2001 this sequence version replaced gi:13699537.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 763224
Center clone name: CITB-H1_2353F3
-----
Summary Statistics
Consensus quality: 127285 bases at least Q40

```

```

Consensus quality: 129762 bases at least Q30
Consensus quality: 130660 bases at least Q20
Estimated insert size: 108000; pulse field gel estimation
Estimated insert size: 131623; sum-of-contigs estimation
Quality coverage: 7.54 in Q20 bases; pulse field gel estimation
Quality coverage: 6.19 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 25177: contig of 25177 bp in length
* 25178 25277: gap of unknown length
* 25278 52649: contig of 27372 bp in length
* 52650 52749: gap of unknown length
* 52750 131823: contig of 79074 bp in length.
FEATURES
            Location/Qualifiers
            source
            1..131823
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /chromosome="5"
            /clone_CTD="2353F3"
            /clone_lib="CalTech human BAC library D"

```

ORIGIN

```

Query Match      58.5%; Score 24; DB 2; Length 131823;
Best Local Similarity 75.0%; Pred. No. 1.5e+02;
Matches 30; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

```

```

Qy 1 TCTCTCTCGACTTTGTCGTAATAATGATAAGATCTAATT 40
    |||||
Db 68042 TCACTGATACTTACTCAGTGAATATATAGATCTAATT 68081
    |||||

```

```

RESULT 11
AC010296/c
LOCUS AC010296      134349 bp DNA linear PRI 27-APR-2002
DEFINITION Homo sapiens chromosome 5 clone CTB-155H23, complete sequence.
ACCESSION AC010296
VERSION AC010296.6 GI:20334544
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 134349)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 134349)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (15-SEP-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Apr 27, 2002 this sequence version replaced gi:15808509.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center

```

RESULT 13

	AC108124	179647 bp	DNA	linear	HTG 25-JAN-2002
LOCUS	AC108124				
DEFINITION	Homo sapiens chromosome 5 clone RP11-56N9, WORKING DRAFT SEQUENCE, 3 unordered pieces.				
ACCESSION	AC108124				
VERSION	AC108124.1	GI:18369970			
KEYWORDS	HTG; HTGS PHASE1; HTGS_DRAFT; HTGS_ACTIVEPIN. Phage phiHmu1				

SOURCE: Homo sapiens (human);
ORGANISM: Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 179647)
Donor: GenBank, Inc.

AUTHORS	DOE JOINT GENOME INSTITUTE.
TITLE	Sequencing of Human Chromosome 5
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 179647)
AUTHORS	DOE Joint Genome Institute.
TITLE	Direct Submission
JOURNAL	submitted (25-JAN-2002) Production Sequencing Facility, DOE Joint

COMMENT
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: <http://www.jgi.doe.gov>
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

```
-----
Project Information
Center Project Name: 428190
Center clone name: RPCI-11 58N9
```

```

-----
Summary Statistics
Consensus quality: 178319 bases at least Q40
Consensus quality: 179012 bases at least Q30
Consensus quality: 179140 bases at least Q20
Estimated insert size: 1794000; agarose-fp estimation
Estimated insert size: 179447; sum-of-contigs estimation
Quality coverage: 3.85 in Q20 bases; agarose-fp estimation
Quality coverage: 4.16 in Q20 bases; sum-of-contigs estimation.
* NOTE: this is a 'working draft' sequence. It currently
* consists of 3 contigs. the true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved. 1993
-----

```

*	1	1933: contig of 1933 bp in length
*	1934	2033: gap of unknown length
*	2034	51443: contig of 49410 bp in length
*	51444	51443: gap of unknown length
*	51544	179647: contig of 138104 bp in length.

```

FEATURES
  source
    Location/Qualifiers
      1..179647
        /organism="Homo sapiens"
        /mol_type="Genomic DNA"
        /db_xref="taxon:9606"
        /chromosome="5"
        /clone="RP11-56N9"
        /clone_lib="RPC1 human BAC library 11"

ORIGIN
  Query Match
    58.5%; Score 24; DB 2; Length 179647;

```

QY 1 TCTCTCTGACTTTGTCGGTAAATTGATAAGATCTTAATTT 40
 Matches 30; Conservative 0; Mismatches 10; Indels 0; Gaps 0
 Best Local Similarity 75.0%; Pred. NO. 1.5e+02;


```

Db      15239  TCACGTGATACCTTACTCAGTCAGATATATAGATCAATTT 15278
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
RESULT 14
AC010233/c
LOCUS   AC010233
DEFINITION Homo sapiens chromosome 5 clone CTC-330K4, WORKING DRAFT SEQUENCE. linear HTG 14-FEB-2001
ACCESSION AC010233
VERSION AC010233.6 GI:12830110
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 189207)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 5
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 189207)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (15-SEP-1999) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT On Feb 14, 2001 this sequence version replaced gi:8575983.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.lqi.doe.gov

```

```

-----
Project Information
Center Project Name: 312143
Center clone name: CIT-HSPC_330K4
-----
Summary Statistics
Consensus quality: 178734 bases at least Q40
Consensus quality: 184726 bases at least Q30
Consensus quality: 186591 bases at least Q20
Estimated insert size: 187000; pulse field gel estimation
Estimated insert size: 187707; sum-of-contigs estimation
Quality coverage: 6.84 in Q20 bases; pulse field gel estimation
Quality coverage: 6.81 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submitter.

```

* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

*	1	22993	contig of 22993 bp in length
*	22994	23093	gap of unknown length
*	23094	25166	contig of 2073 bp in length
*	25167	25266	gap of unknown length
*	25267	100726	contig of 75460 bp in length
*	100727	100826	gap of unknown length
*	100827	135879	contig of 35053 bp in length
*	135880	135979	gap of unknown length
*	135980	137446	contig of 1767 bp in length
*	137447	137846	gap of unknown length
*	137847	188207	contig of 50361 bp in length.

[illegible]

```
1. 16820 /
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTC-330K4"
/clone_lib="Caltech human
```

ORIGIN

```

Query Match          59.5%;   Score 24;   DB 2;   Length 188207;
Best Local Similarity 75.0%;   Pred. No. 1.5e+02;
Matches 30; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 1 TCCTCTCGACTTGTCCGTAATAATTGATAAGATCTAAATTT 40
    ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 85199 TCCTGTAACCTACTCAGTGAATATATAAGATCTAAATTT 85160

RESULT 15
AR394374/c
LOCUS AR394374 1347 bp DNA linear PAT 18-DEC-2003
DEFINITION Sequence 389 from patent US 6617156.
ACCESSION AR394374
VERSION AR394374.1 GI:40122001
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
UNclassified.
REFERENCE 1 (bases 1 to 1347)
AUTHORS Doucette-Stamm,L.A. and Bush,D.
TITLE Nucleic acid and amino acid sequences relating to Enterococcus
faecalis for diagnostics and therapeutics
JOURNAL Patent: US 6617156-A 389 09-SEP-2003;
FEATURES
    source
        1. .1347
            /organism="unknown"
            /mol_type="genomic DNA"

ORIGIN

Query Match          58.0%;   Score 23.8;   DB 6;   Length 1347;
Best Local Similarity 80.0%;   Pred. No. 2.1e+02;
Matches 28; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 3 TCCTCCTGACTTTGTCGCTAAATGATAAGATCTAA 37
    ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1151 TCCTCCTTCATTCGCTAAATGATAATCTCTAA 1117

Search completed: November 3, 2004, 07:43:05
Job time : 261.809 secs

```

Search completed: November 3, 2004, 07:43:05
Job time : 261.809 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 3, 2004, 04:44:23 ; Search time 687.273 Seconds
(without alignments)
2566.384 Million cell updates/sec

Title: US-09-933-267a-1_COPY_306168_306503
Perfect score: 336
Sequence: 1 ggatacgaagaccgaaga.....ctggcggaagagggtgccag 336

Scoring table: IDENTITY NUC
Gapop 10_0 , Gapext 1.0

Searched: 4134886 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_23Sep04:*

- 1: Geneseqn1980s:*
- 2: Geneseqn1990s:*
- 3: Geneseqn2000s:*
- 4: Geneseqn2001as:*
- 5: Geneseqn2001bs:*
- 6: Geneseqn2002as:*
- 7: Geneseqn2002bs:*
- 8: Geneseqn2003as:*
- 9: Geneseqn2003bs:*
- 10: Geneseqn2003cs:*
- 11: Geneseqn2003ds:*
- 12: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	336	100.0	1044	6	ABZ23390	Nucleotid
2	336	100.0	1044	6	ABZ23391	Reverse c
3	336	100.0	1223	4	AAF29924	Human est
4	336	100.0	1237	6	ABK89697	Oestrogen
5	336	100.0	1374	4	AAF29925	Human est
6	336	100.0	1380	4	AAC86919	Nucleotid
7	336	100.0	1770	10	ADB99351	Fusion pr
8	336	100.0	1785	10	AAD64536	Human wil
9	336	100.0	1788	6	ABL57497	Human nuc
10	336	100.0	1803	10	ADB99349	Fusion pr
11	336	100.0	2028	10	AAD64590	Human ER3
12	336	100.0	2031	10	AAD64588	Human ER3
13	336	100.0	2092	1	AAW70880	Human ER3
14	336	100.0	2092	4	AAC86920	cDNA enco
15	336	100.0	2092	6	ABK89700	Oestrogen
16	336	100.0	2092	6	ABL51900	Human oes
17	336	100.0	2092	8	AAD53909	Human wil
18	336	100.0	2092	10	ADC09969	Human oes
19	336	100.0	2106	6	ABK89699	Oestrogen
20	336	100.0	2178	10	AAD64589	Human ER3
21	336	100.0	2181	10	AAD64587	Human ER3

22	336	100.0	2220	6	ABS56328	Human oes
23	336	100.0	2220	6	ABA01104	Fused ER
24	336	100.0	2220	10	ADC51161	Human oes
25	336	100.0	2322	2	AAK60628	SSR-LBD f
26	336	100.0	2532	10	ADC51504	Human GST
27	336	100.0	6450	2	AAZ23433	Human est
28	336	100.0	6450	6	ABK89695	Oestrogen
29	336	100.0	6450	6	ABK89703	Oestrogen
30	336	100.0	6450	6	ABV94431	Breast ca
31	336	100.0	6450	6	ABZ23389	Reverse c
32	336	100.0	6450	6	ABZ23388	Nucleotid
33	336	100.0	6450	8	ABX12085	cDNA enco
34	336	100.0	6450	8	ACC50140	Breast oes
35	336	100.0	6450	8	ABX93797	Human oes
36	336	100.0	6450	10	ADB81383	DNA seque
37	336	100.0	6450	10	ADD25499	Binding d
38	336	100.0	6450	10	ADE12135	Human oes
39	336	100.0	6450	10	ADF76402	Novel hum
40	336	100.0	6450	10	ADG89355	Cancer de
41	336	100.0	6450	12	ADF83093	Human tes
42	336	100.0	6450	12	ADF72329	Human and
43	336	100.0	6450	12	ADP07306	Human ESR
44	336	100.0	6450	12	ADP05662	Human nuc
45	336	100.0	6468	8	ACC46160	Human dit

ALIGNMENTS

RESULT 1
ABZ23390
ID ABZ23390 standard; cDNA; 1044 BP.
XX
AC ABZ23390;
XX
DT 07-APR-2003 (first entry)
XX
DE Nucleotide sequence of human oestrogen receptor alpha bait protein.
XX
KW Human; oestrogen receptor alpha; ER-alpha; CF16; CF17; CF18; CF19; CF40;
KW CF41; CF42; CF43; cofactor; osteoporosis; bone disease; reproduction;
KW hormonal dysfunction; cancer; cardiovascular disease; atherosclerosis;
KW hot flush; mood change; Alzheimer's disease; gene; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..1044
FT /tag= a
FT /product= "oestrogen receptor alpha bait"
XX
PW WO200270699-A2.
XX
PD 12-SEP-2002.
XX
PF 28-FEB-2002; 2002WO-EP002189.
XX
PR 01-MAR-2001; 2001EP-00105062.
XX
PA (LION-) LION BIOSCIENCE AG.
XX
PI Albers M, Ellwanger S, Loeser E, Koegl M;
XX
DR WPI; 2002-713451/77.
XX
DR P-PSDB; ABP70164.
XX
PT New cofactors of estrogen receptor alpha, designated as CF16, CF17, CF18,
PT CF19, CF40, CF41, CF42 and/or CF43, useful for screening of compounds for
PT treating osteoporosis, hormonal dysfunctions, cancer or cardiovascular
PT diseases.
XX
PS Disclosure; Page 108-109; 111pp; English.
XX

CC The present sequence encodes a human oestrogen receptor alpha (ER-alpha)
CC bait protein. The specification describes cofactors of ER-alpha.
CC designated CF16, CF17, CF18, CF19, CF40, CF41, CF42, and CF43. The
CC cofactor polypeptides and nucleic acid molecules are useful for screening
CC for compounds for treating osteoporosis and other bone diseases, failures
CC in reproductive functions or hormonal dysfunctions, cancer or
CC cardiovascular diseases such as atherosclerosis, and in preventing hot
CC flushes, mood changes and Alzheimer's disease. The CF proteins are also
CC useful for screening for ligands of the ER alpha. The nucleic acid
CC sequences are useful for making vectors and CF polypeptides, transforming
CC host cells, as research tools for developing nucleic acid probes, and for
CC developing analytical tools such as antisense oligonucleotides
XX
SQ Sequence 1044 BP; 252 A; 272 C; 295 G; 225 T; 0 U; 0 Other;

Query Match 100.0%; Score 336; DB 6; Length 1044;
Best Local Similarity 100.0%; Pred. No. 9.9e-92;
Matches 336; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GGATAGCAAAAGACCGAAGAGGAGGAGATGTTGAACACACAGCGCCGAGAGATGATG 60
Db 17 GGATAGCAAAAGACCGAAGAGGAGGAGATGTTGAACACACAGCGCCGAGAGATGATG 76
QY 61 GCGAGGCGAGGAGTGAAGTGGGTCTGCTGGAGACATGAGAGCTGCCAACCTTTGGCCAA 120
Db 77 GCGAGGCGAGGAGTGAAGTGGGTCTGCTGGAGACATGAGAGCTGCCAACCTTTGGCCAA 136
QY 121 GCCCGCTCATGATCAACACGCTCTAAGAAGAACAGCGCTGGCTTGTCCCTGACGGCGGACC 180
Db 137 GCCCGCTCATGATCAACACGCTCTAAGAAGAACAGCGCTGGCTTGTCCCTGACGGCGGACC 196
QY 181 AGATGGTCACTGCTTGTGGATGCTGAGCGCCCGCCATCTATTCGAGTATGATCCTA 240
Db 197 AGATGGTCACTGCTTGTGGATGCTGAGCGCCCGCCATCTATTCGAGTATGATCCTA 256
QY 241 CCAGACCTTCAGTGAAGCTTCGATGATGGCTTACTGACCAACCTGGCAGACAGGGAGC 300
Db 257 CCAGACCTTCAGTGAAGCTTCGATGATGGCTTACTGACCAACCTGGCAGACAGGGAGC 316
QY 301 TGGTTCACTGATCAACTGGGCGAAGAGGGTGCCAG 336
Db 317 TGGTTCACTGATCAACTGGGCGAAGAGGGTGCCAG 352

RESULT 2
ABZ23391/c
ID ABZ23391 standard; cDNA; 1044 BP.

XX AC ABZ23391;

XX DT 07-APR-2003 (first entry)

XX Reverse complement of human oestrogen receptor alpha bait protein cDNA.

XX Human; oestrogen receptor alpha; ER-alpha; CF16; CF17; CF18; CF19; CF40;
XX CF41; CF42; CF43; cofactor; osteoporosis; bone disease; reproduction;
XX hormonal dysfunction; cancer; cardiovascular disease; atherosclerosis;
XX hot flush; mood change; Alzheimer's disease; ss.

XX Homo sapiens.

XX WO200270699-A2.

XX 12-SEP-2002.

XX 28-FEB-2002; 2002WO-EP002189.

XX 01-MAR-2001; 2001EP-00105062.

XX (LION-) LION BIOSCIENCE AG.

XX Albers M, Ellwanger S, Loeser E, Koegl M;

XX

DR WPI; 2002-713451/77.

XX New cofactors of estrogen receptor alpha, designated as CF16, CF17, CF18,
XX CF19, CF40, CF41, CF42 and/or CF43, useful for screening of compounds for
XX treating osteoporosis, hormonal dysfunctions, cancer or cardiovascular
XX diseases.

XX Disclosure; Page 109-110; 111pp; English.

XX The present sequence represents the reverse complement of an oestrogen
XX receptor alpha (ER-alpha) bait protein. The specification describes
XX cofactors of ER-alpha, designated CF16, CF17, CF18, CF19, CF40, CF41,
XX CF42, and CF43. The cofactor polypeptides and nucleic acid molecules are
XX useful for screening for compounds for treating osteoporosis and other
XX bone diseases, failures in reproductive functions or hormonal
XX dysfunctions, cancer or cardiovascular diseases such as atherosclerosis,
XX and in preventing hot flushes, mood changes and Alzheimer's disease. The
XX CF proteins are also useful for screening for ligands of the ER alpha.
XX The nucleic acid sequences are useful for making vectors and CF
XX polypeptides, transforming host cells, as research tools for developing
XX nucleic acid probes, and for developing analytical tools such as
XX antisense oligonucleotides

XX SQ Sequence 1044 BP; 225 A; 295 C; 272 G; 252 T; 0 U; 0 Other;

Query Match 100.0%; Score 336; DB 6; Length 1044;
Best Local Similarity 100.0%; Pred. No. 9.9e-92;
Matches 336; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGATAGCAAAAGACCGAAGAGGAGGAGATGTTGAACACACAGCGCCGAGAGATGATG 60
Db 1028 GGATAGCAAAAGACCGAAGAGGAGGAGATGTTGAACACACAGCGCCGAGAGATGATG 969
QY 61 GCGAGGCGAGGAGTGAAGTGGGTCTGCTGGAGACATGAGAGCTGCCAACCTTTGGCCAA 120
Db 968 GCGAGGCGAGGAGTGAAGTGGGTCTGCTGGAGACATGAGAGCTGCCAACCTTTGGCCAA 909
QY 121 GCCCGCTCATGATCAACACGCTCTAAGAAGAACAGCGCTGGCTTGTCCCTGACGGCGGACC 180
Db 908 GCCCGCTCATGATCAACACGCTCTAAGAAGAACAGCGCTGGCTTGTCCCTGACGGCGGACC 849
QY 181 AGATGGTCACTGCTTGTGGATGCTGAGCGCCCGCCATCTATTCGAGTATGATCCTA 240
Db 848 AGATGGTCACTGCTTGTGGATGCTGAGCGCCCGCCATCTATTCGAGTATGATCCTA 789
QY 241 CCAGACCTTCAGTGAAGCTTCGATGATGGCTTACTGACCAACCTGGCAGACAGGGAGC 300
Db 788 CCAGACCTTCAGTGAAGCTTCGATGATGGCTTACTGACCAACCTGGCAGACAGGGAGC 729
QY 301 TGGTTCACTGATCAACTGGGCGAAGAGGGTGCCAG 336
Db 728 TGGTTCACTGATCAACTGGGCGAAGAGGGTGCCAG 693

RESULT 3

AAF29924

ID AAF29924 standard; DNA; 1223 BP.

XX AC AAF29924;

XX DT 04-APR-2001 (first entry)

XX Human estrogen receptor alpha isoform #1 DNA.

XX Human; estrogen receptor alpha; cancer; osteoporosis; bone; Alzheimer's;
XX cardiovascular; ds.

XX Homo sapiens.

XX WO200100823-A1.

XX 04-JAN-2001.

XX

PF 27-JUN-2000; 2000WO-EP005981.
XX
PR 29-JUN-1999; 99IT-MI001433.
XX
PA (BUMO-) EURO MOLECULAR BIOLOGY LAB.
XX
PI Gannon F, Dengler S, Flouriot G;
XX
DR WPI; 2001-137955/14.
XX
XX
PT Novel isoforms of human estrogen receptor alpha useful for preparing
PT therapeutic agents for treating cancer, osteoporosis, Alzheimer's disease
PT and cardiovascular diseases.
XX
PS Claim 1; Pag e44; 53pp; English.
XX
XX The present invention relates to a human estrogen receptor (hER)-alpha
XX isoform. Molecules which modulate the activity of the estrogen receptor
XX are useful for the preparation of therapeutic agents for treating cancer,
XX osteoporosis and other bone disorders, Alzheimer's disease and
XX cardiovascular diseases
XX
SQ Sequence 1223 BP; 303 A; 316 C; 341 G; 263 T; 0 U; 0 Other;
Query Match 100.0%; Score 336; DB 4; Length 1223;
Best Local Similarity 100.0%; Pred. No. 1.1e-91;
Matches 336; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GGATACGAAAGACCGAAGAGGAGGAGGAGATGTTGAAACACAAAGCCCGAGAGATGATG 60
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 196 GGATACGAAAGACCGAAGAGGAGGAGGAGATGTTGAAACACAAAGCCCGAGAGATGATG 255
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 61 GGGAGGCGAGGGGTGAAGTGGGTCTGCTGGAGACATGAGCTGCCAACCTTTGGCCAA 120
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 256 GGGAGGCGAGGGGTGAAGTGGGTCTGCTGGAGACATGAGCTGCCAACCTTTGGCCAA 315
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 121 GCCCGCTCATGATCAAAACGCTTAAGAAAGACAGCTGCGCTTGTCCCTGACGCGGACC 180
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 316 GCCCGCTCATGATCAAAACGCTTAAGAAAGACAGCTGCGCTTGTCCCTGACGCGGACC 375
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 181 AGATGTCAGTGCCTTGTGGATGCTGAGCCCCCATCTACTTATTCGGAGTATGATCCTA 240
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 376 AGATGTCAGTGCCTTGTGGATGCTGAGCCCCCATCTACTTATTCGGAGTATGATCCTA 435
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 241 CCAGACCCCTTCAGTGAAGCTTCGATGATGGCTTACTGACCAACCTGGCAGACAGGGAGC 300
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 436 CCAGACCCCTTCAGTGAAGCTTCGATGATGGCTTACTGACCAACCTGGCAGACAGGGAGC 495
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 301 TGGTTACATGATCAACTGGGCGAAGAGGGTGCCAG 336
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 496 TGGTTACATGATCAACTGGGCGAAGAGGGTGCCAG 531
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
RESULT 4
ABK89697
ID ABK89697 standard; DNA; 1237 BP.
XX
AC ABK89697;
XX
DT 05-NOV-2002 (first entry)
XX
DE Oestrogen receptor alpha nucleic acid comprising A908G mutation, #3.
XX
KW Oestrogen receptor alpha; breast cancer; pre-malignant lesion;
KW invasive breast cancer; A908G oestrogen receptor alpha transition;
KW oestrogen receptor alpha K303R substitution; human; gene; ds.
XX
OS Homo sapiens.
XX
PN WO200257283-A1.
XX
PD 25-JUL-2002.
XX

PF 16-JAN-2002; 2002WO-US004982.
XX
PR 19-JAN-2001; 2001US-0262990P.
PR 09-JUL-2001; 2001US-0304018P.
XX
PA (BAYU) BAYLOR COLLEGE MEDICINE.
XX
XX Fuqua S, O'connell P, Allred DC, Hopp TA;
PI WPI; 2002-590711/63.
XX
DR New isolated estrogen receptor alpha with A908G mutation or K303R
XX substitution, useful as diagnostic marker in breast tissue such as pre-
PT malignant lesions for the development of breast cancer, particularly
PT invasive breast cancer.
XX
XX Claim 1; Page; 133pp; English.
XX
XX The invention relates to an isolated oestrogen receptor alpha nucleic
XX acid sequence comprising an A908G mutation, or an amino acid sequence
XX comprising a K303R substitution. Also described are methods for detecting
XX susceptibility to development of breast cancer or invasive breast cancer
XX in an individual, for diagnosing breast cancer in an individual; and for
XX screening for a modulator of an oestrogen receptor alpha polypeptide
XX comprising a K303R substitution. The oestrogen receptor alpha is useful
XX as a diagnostic marker in breast tissue such as pre-malignant lesions
XX for the development of breast cancer, particularly invasive breast
XX cancer. The methods are useful for determining susceptibility to
XX development of breast cancer, for diagnosing, preventing or treating
XX breast cancer. Transgenic mice may be used for screening and identifying
XX agents that interact with the oestrogen receptor alpha, or affect breast
XX tissue health. The A908G oestrogen receptor alpha transition is
XX frequently present in pre-malignant lesions of the breast and can occur
XX in the adjacent normal-appearing breast epithelium. The present sequence
XX represents an oestrogen receptor alpha nucleic acid which may comprise
XX the A908G mutation. Note: The present sequence is not shown in the
XX specification but was obtained by the indexer from GenBank using the
XX accession number given in the specification
SQ Sequence 1237 BP; 369 A; 218 C; 286 G; 364 T; 0 U; 0 Other;
Query Match 100.0%; Score 336; DB 6; Length 1237;
Best Local Similarity 100.0%; Pred. No. 1.1e-91;
Matches 336; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GGATACGAAAGACCGAAGAGGAGGAGGAGATGTTGAAACACAAAGCCCGAGAGATGATG 60
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 350 GGATACGAAAGACCGAAGAGGAGGAGGAGATGTTGAAACACAAAGCCCGAGAGATGATG 409
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 61 GGGAGGCGAGGGGTGAAGTGGGTCTGCTGGAGACATGAGAGCTGCCAACCTTTGGCCAA 120
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 410 GGGAGGCGAGGGGTGAAGTGGGTCTGCTGGAGACATGAGAGCTGCCAACCTTTGGCCAA 469
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 121 GCCCGCTCATGATCAAAACGCTTAAGAAAGACAGCTGCGCTTGTCCCTGACGCGGACC 180
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 470 GCCCGCTCATGATCAAAACGCTTAAGAAAGACAGCTGCGCTTGTCCCTGACGCGGACC 529
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 181 AGATGTCAGTGCCTTGTGGATGCTGAGCCCCCATCTACTTATTCGGAGTATGATCCTA 240
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 530 AGATGTCAGTGCCTTGTGGATGCTGAGCCCCCATCTACTTATTCGGAGTATGATCCTA 589
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 241 CCAGACCCCTTCAGTGAAGCTTCGATGATGGCTTACTGACCAACCTGGCAGACAGGGAGC 300
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 590 CCAGACCCCTTCAGTGAAGCTTCGATGATGGCTTACTGACCAACCTGGCAGACAGGGAGC 649
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 301 TGGTTACATGATCAACTGGGCGAAGAGGGTGCCAG 336
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 650 TGGTTACATGATCAACTGGGCGAAGAGGGTGCCAG 685
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
RESULT 5
AAF29925
ID AAF29925 standard; DNA; 1374 BP.


```

RESULT 7
ADB99351
ID ADB99351 standard; DNA; 1770 BP.
XX AC
XX ADB99351;
XX DT 04-DEC-2003 (first entry)
XX DE
XX FUSION protein GAL4-AD-ERADAB DNA sequence related to ER constructs.
XX KW Yeast two hybrid assay; oestrogen receptor-alpha receptor;
XX KW ER-alpha receptor; oestrogen receptor; ER; 17beta-oestradiol;
XX KW nuclear receptor; steroid receptor; transactivation factor; TAF-1;
XX KW Yeast two hybrid system; GAL4-AD-ERADAB; fusion protein; gene; ds; human;
XX OS
XX OS Chimeric.
XX OS Unidentified.
XX OS Homo sapiens.
XX OS Ascomycota.
XX FH
XX Key Location/Qualifiers
XX CDS 1..1770
XX /tag= b
XX /product= "GAL4-AD-ERADAB fusion protein"
XX misc_feature 1..417
XX /tag= a
XX /note= "This region encodes the GAL4 activation domain
XX from yeast"
XX misc_feature 418..1770
XX /tag= c
XX /note= "This region encodes the truncated human ERalpha"
XX PN
XX WO2003070975-A2.
XX PD
XX 28-AUG-2003.
XX XX
XX 21-FEB-2003; 2003WO-EP050029.
XX XX
XX 25-FEB-2002; 2002EP-00075788.
XX XX
XX (ALKU ) AKZO NOBEL NV.
XX XX
XX Verdonk G, Dijkema R, Schoonen WGEJ;
XX WPI; 2003-663859/62.
XX P-PSDB; ADB99352.
XX XX
XX New gene construct comprising the code for a truncated estrogen receptor
XX (ER)-alpha receptor having regions A and B that are absent, useful for
XX testing compound for their influence of dimerization.
XX XX
XX Example 3; Page 17-19; 27pp; English.
XX CC
XX This invention relates to a novel gene construct for use in a yeast two
XX hybrid assay, comprising the code for a truncated oestrogen receptor-
XX alpha (ER-alpha) receptor having regions A and B that are absent. An
XX oestrogen receptor (ER) is a protein that binds the natural hormone
XX 17beta-oestradiol and belongs to the family of nuclear receptors and more
XX specifically to the subfamily of steroid receptors. ER-alpha contains
XX transactivation factor (TAF-1) regions A and B which are absent in the
XX truncated ER-alpha of the current invention. This provides the advantage
XX that the false negative or basal activation in the yeast two hybrid
XX system is reduced in comparison to the full length construct. The gene
XX construct may be useful for use in the yeast two hybrid assay for testing
XX compound for their influence of dimerisation and/or the transcription
XX promoting function of dimerisation and selecting a compound out of the
XX compounds which is selectively active on the ER-alpha/ER-beta
XX heterodimer. The present sequence is the DNA sequence which encodes a
XX GAL4-AD-ERADAB fusion protein used in the exemplification of the
XX invention.
XX SQ
Sequence 1770 BP; 489 A; 438 C; 452 G; 391 T; 0 U; 0 Other;

```

```

Query Match 100.0%; Score 336; DB 10; Length 1770;
Best Local Similarity 100.0%; Pred. No. 1.2e-91;
Matches 336; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGATACGAAAGACCCGAGAGGAGGAGAGTCTTGAACACACAGGCCAGAGAGATGATG 60
DB 743 GGATACGAAAGACCCGAGAGGAGGAGAGTCTTGAACACACAGGCCAGAGAGATGATG 802
QY 61 GGGAGGGCAGGGGTGAAGTGGGGTCTGTGGAGACATGAGAGCTGCCAACCTTTGGCCAA 120
DB 803 GGGAGGGCAGGGGTGAAGTGGGGTCTGTGGAGACATGAGAGCTGCCAACCTTTGGCCAA 862
QY 121 GCCCGCTCATGATCAACAGCGCTCTAAGAAAGACAGCGCTTGCCTTGCCTGACGGCGGACC 180
DB 863 GCCCGCTCATGATCAACAGCGCTCTAAGAAAGACAGCGCTTGCCTTGCCTGACGGCGGACC 922
QY 181 AGATGTCAGTGCCTTGTGGATGCTGAGCCCCCATACTCTATTCCGAGTATGATCCTA 240
DB 923 AGATGTCAGTGCCTTGTGGATGCTGAGCCCCCATACTCTATTCCGAGTATGATCCTA 982
QY 241 CCAGACCTTCAGTGAAGCTTCGATGATGGGTCTTACTGACCAACCTGCGACAGCGGAGC 300
DB 983 CCAGACCTTCAGTGAAGCTTCGATGATGGGTCTTACTGACCAACCTGCGACAGCGGAGC 1042
QY 301 TGGTTCACATGATCAACTGGGCGAAGAGGGTGCCAG 336
DB 1043 TGGTTCACATGATCAACTGGGCGAAGAGGGTGCCAG 1078

RESULT 8
AAD64536
ID AAD64536 standard; DNA; 1785 BP.
XX AC
XX AAD64536;
XX DT 12-FEB-2004 (first entry)
XX DE
XX Human wild-type oestrogen receptor (ER) alpha DNA.
XX KW Tamoxifen activated system; TAS; 4-hydroxytamoxifen; OHT; gene therapy;
XX KW oestrogen receptor; ER; fundamental research; transgenic; biomedical;
XX KW disease model; breast cancer; human; gene; ds.
XX OS
XX Homo sapiens.
XX FH
XX Key Location/Qualifiers
XX CDS 1..1785
XX /tag= a
XX /product= "Human ERalpha protein"
XX /note= "CDS does not include stop codon"
XX /partial
XX XX
XX US2003199022-A1.
XX PD
XX 23-OCT-2003.
XX XX
XX 11-MAR-2002; 2002US-00095373.
XX XX
XX 11-MAR-2002; 2002US-00095373.
XX XX
XX (UNII ) UNIV ILLINOIS FOUND.
XX XX
XX Mao C, Shapiro DJ;
XX WPI; 2003-852787/79.
XX XX
XX New isolated polynucleotide useful for gene therapy applications (e.g. in
XX treating breast cancer), in vivo and in vitro gene expression, in
XX producing bioactive or toxic polypeptides, in research, or in producing
XX transgenic animals.
XX PS
Disclosure; SEQ ID NO 1; Opp; English.

```

XX The invention relates to tamoxifen activated system (TAS) and 4-
CC hydroxytamoxifen (OHT) activated system for regulated production of
CC proteins in eukaryotic cells. TAS includes mutant oestrogen receptors
CC (ERS) and chimeras thereof. The invention is useful in gene therapy (e.g.
CC in treating breast cancer) or in vivo and in vitro gene expression, in
CC producing bioactive, toxic, recombinant polypeptides in mammalian cells,
CC in biomedical and fundamental research, or in producing transgenic
CC animals. The transgenic animals may be used as disease models, in
CC studying the function and/or activity of a polypeptide, or in identifying
CC and/or evaluating modulators of a polypeptide activity. The present
CC sequence is human wild-type ERalpha DNA used in TAS
XX
SQ Sequence 1785 BP; 417 A; 515 C; 506 G; 347 T; 0 U; 0 Other;

Query Match 100.0%; Score 336; DB 10; Length 1785;
Best Local Similarity 100.0%; Pred. No. 1.2e-91;
Matches 336; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GGATACGAAAGACCGAAGAGGGGAGGAGATGTTGAAACACACAGCCGAGAGATGATG 60
DB 761 GGATACGAAAGACCGAAGAGGGGAGGAGATGTTGAAACACACAGCCGAGAGATGATG 820
QY 61 GGGAGGCGAGGGGTGAAGTGGGTCTCTGGAGACATGAGAGCTGCCAACCTTTGGCCAA 120
DB 821 GGGAGGCGAGGGGTGAAGTGGGTCTCTGGAGACATGAGAGCTGCCAACCTTTGGCCAA 880
QY 121 GCCCGCTCATGATCAACCGCTCTAAGAGAACAGCCTGGCTTGTCCCTGACGGCGACC 180
DB 881 GCCCGCTCATGATCAACCGCTCTAAGAGAACAGCCTGGCTTGTCCCTGACGGCGACC 940
QY 181 AGATGGTCAGTGCTTGTGGATGCTGAGCCGCCCTACTCTATTCGGATGATGCTCTA 240
DB 941 AGATGGTCAGTGCTTGTGGATGCTGAGCCGCCCTACTCTATTCGGATGATGCTCTA 1000
QY 241 CCAGACCCCTTCAGTGAAGCTTCGATGATGGCTTACTGACCAACCTGGCAGACAGGAGC 300
DB 1001 CCAGACCCCTTCAGTGAAGCTTCGATGATGGCTTACTGACCAACCTGGCAGACAGGAGC 1060
QY 301 TGGTTACATGATCAACTGGCGAAGAGGGTGCAG 336
DB 1061 TGGTTACATGATCAACTGGCGAAGAGGGTGCAG 1096

RESULT 9
ABL57497
ID ABL57497 standard; cDNA; 1788 BP.
XX ABL57497;
XX 22-AUG-2002 (first entry)
XX Human nuclear oestrogen receptor alpha coding sequence.
XX Oestrogen; receptor; human; transgenic mouse; cytostatic;
XX antiinflammatory; antidiabetic; endocrine; anorectic; hepatotropic; gene;
XX ss.
XX Homo sapiens.
XX Key Location/Qualifiers
XX CDS 1..1788
XX /*tag= a
XX /product= "Nuclear oestrogen receptor"
XX W0200228175-A2.
XX 11-APR-2002.
XX 28-SEP-2001; 2001WO-IB002246.
XX 03-OCT-2000; 2000FR-00012570.
XX 11-MAY-2001; 2001US-00853033.

XX (ADER-) ADEREGEM ASSOC DEV RECH EN GENETIQ.
XX Chambon P, Metzger D;
XX WPI; 2002-463217/49.
XX P-P8DB; ABB76378.
XX A transgenic mouse, useful in screening for medicaments for the treatment
XX of e.g., diabetes or skin cancers, comprises a fusion protein between a
XX recombinase Cre, and a modified ligand binding domain of the nuclear
XX estrogen receptor alpha.
XX
XX Disclosure; Page 128-130; 149pp; English.
XX
XX The present sequence is the coding sequence of the human nuclear
XX oestrogen receptor alpha (I). The invention relates to a non-human
XX metazoan organism, especially a transgenic mouse, characterised in that
XX at least one cell comprises: (i) a fusion protein formed from a
XX recombinase (Cre), a hinge region (preferably derived from human (I)),
XX and a modified ligand binding domain (LBD) of a nuclear oestrogen
XX receptor, such as (I), its fragment or variant; and (ii) one or more
XX genes or DNA sequences of interest belonging to the genome of the
XX organism, into which one or more recognition sites of the recombinase
XX protein are inserted. The (I) LBD domain is preferably modified by a
XX G521R, G400V, or M543A/L544A mutation. The fusion protein has negligible,
XX or even zero, recombinase activity in the presence of a natural ligand
XX such as oestradiol, but recombinase activity is induced by a small
XX quantity of a synthetic ligand that has antioestrogenic activity, e.g.
XX tamoxifen or 4-hydroxytamoxifen. The metazoan organism or its cells, such
XX as epidermal cells, hepatocytes or adipocytes, are useful in carrying out
XX a spatiotemporally controlled site-specific recombination of a DNA
XX sequence of interest in its natural chromatin environment. It is also
XX used in screening of medicaments for pathological conditions associated
XX with an alteration of the expression and/or function of the DNA sequence
XX of interest, such as skin cancer, inflammation, diabetes, alopecia,
XX obesity, or in promoting hepatic regeneration
XX
SQ Sequence 1788 BP; 418 A; 515 C; 508 G; 347 T; 0 U; 0 Other;

Query Match 100.0%; Score 336; DB 6; Length 1788;
Best Local Similarity 100.0%; Pred. No. 1.2e-91;
Matches 336; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GGATACGAAAGACCGAAGAGGGGAGGAGATGTTGAAACACACAGCCGAGAGATGATG 60
DB 761 GGATACGAAAGACCGAAGAGGGGAGGAGATGTTGAAACACACAGCCGAGAGATGATG 820
QY 61 GGGAGGCGAGGGGTGAAGTGGGTCTCTGGAGACATGAGAGCTGCCAACCTTTGGCCAA 120
DB 821 GGGAGGCGAGGGGTGAAGTGGGTCTCTGGAGACATGAGAGCTGCCAACCTTTGGCCAA 880
QY 121 GCCCGCTCATGATCAACCGCTCTAAGAGAACAGCCTGGCTTGTCCCTGACGGCGACC 180
DB 881 GCCCGCTCATGATCAACCGCTCTAAGAGAACAGCCTGGCTTGTCCCTGACGGCGACC 940
QY 181 AGATGGTCAGTGCTTGTGGATGCTGAGCCGCCCTACTCTATTCGGATGATGCTCTA 240
DB 941 AGATGGTCAGTGCTTGTGGATGCTGAGCCGCCCTACTCTATTCGGATGATGCTCTA 1000
QY 241 CCAGACCCCTTCAGTGAAGCTTCGATGATGGCTTACTGACCAACCTGGCAGACAGGAGC 300
DB 1001 CCAGACCCCTTCAGTGAAGCTTCGATGATGGCTTACTGACCAACCTGGCAGACAGGAGC 1060
QY 301 TGGTTACATGATCAACTGGCGAAGAGGGTGCAG 336
DB 1061 TGGTTACATGATCAACTGGCGAAGAGGGTGCAG 1096

RESULT 10
ADB99349
ID ADB99349 standard; DNA; 1803 BP.
XX

ADB99349;
 04-DEC-2003 (first entry)
 Fusion protein GAL4-EDB-ERadAB DNA sequence related to ER constructs.
 yeast two hybrid assay; oestrogen receptor-alpha receptor;
 ER-alpha receptor; oestrogen receptor; ER; 17beta-oestradiol;
 nuclear receptor; steroid receptor; transactivation factor; TAF-1;
 yeast two hybrid system; GAL4-EDB-ERadAB; fusion protein; gene; ds;
 human; yeast.
 Chimeric.
 Unidentified.
 OS Homo sapiens.
 OS Ascomycota.
 XX
 FH Key Location/Qualifiers
 CDS 1..1803
 /tag= b
 /product= "GAL4-DBD-ERadAB fusion protein"
 misc_feature 1..450
 /tag= a
 /note= "This region encodes the GAL4 DNA binding domain
 from Yeast"
 451..1803
 /tag= c
 misc_feature
 /note= "This region encodes the truncated human ERalpha"
 WO2003070975-A2.
 28-AUG-2003.
 21-FEB-2003; 2003WO-EP050029.
 25-FEB-2002; 2002EP-00075788.
 (ALKU) AKZO NOBEL NV.
 Verdonk G, Dijkema R, Schoonen WGEJ;
 WPI; 2003-663859/62.
 P-PSDB; ADB99350.
 New gene construct comprising the code for a truncated estrogen receptor
 (ER)-alpha receptor having regions A and B that are absent, useful for
 testing compound for their influence of dimerization.
 Example 2; Page 14-16; 27pp; English.
 This invention relates to a novel gene construct for use in a yeast two
 hybrid assay, comprising the code for a truncated oestrogen receptor-
 alpha (ER-alpha) receptor having regions A and B that are absent. An
 oestrogen receptor (ER) is a protein that binds the natural hormone
 17beta-oestradiol and belongs to the family of nuclear receptors and more
 specifically to the subfamily of steroid receptors. ER-alpha contains
 transactivation factor (TAF-1) regions A and B which are absent in the
 truncated ER-alpha of the current invention. This provides the advantage
 that the false negative or basal activation in the yeast two hybrid
 system is reduced in comparison to the full length construct. The gene
 construct may be useful for use in the yeast two hybrid assay for testing
 compound for their influence of dimerisation and/or the transcription
 promoting function of dimerisation and selecting a compound out of the
 compounds which is selectively active on the ER-alpha/ER-beta
 heterodimer. The present sequence is the DNA sequence which encodes a
 GAL4-EDB-ERadAB fusion protein used in the exemplification of the
 invention.
 Sequence 1803 BP; 495 A; 432 C; 479 G; 397 T; 0 U; 0 Other;
 Query Match 100.0%; Score 336; DB 10; Length 1803;
 Best Local Similarity 100.0%; Pred. No. 1.2e-91;
 Matches 336; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 XX
 SQ

QY 1 GGATACGAAACACCGAGAGAGGAGGAGATGTTGAAACACACAGCCGACAGAGATGATG 60
 DB 776 GGATACGAAACACCGAGAGAGGAGGAGATGTTGAAACACACAGCCGACAGAGATGATG 835
 QY 61 GCGAGGCGAGGGGTGAAGTGGGGTCTGCTGGAGACATGAGAGCTGCCAACCTTTGGCCAA 120
 DB 836 GCGAGGCGAGGGGTGAAGTGGGGTCTGCTGGAGACATGAGAGCTGCCAACCTTTGGCCAA 895
 QY 121 GCCCGCTCATGATCAAAACGCTCTAAGAAGAAACAGCGCTGGCCCTTGTCCCTGAGCGCGGACC 180
 DB 896 GCCCGCTCATGATCAAAACGCTCTAAGAAGAAACAGCGCTGGCCCTTGTCCCTGAGCGCGGACC 955
 QY 181 AGATGCTCAGTGCCTTGTGGATGCTGAGCCCCCATCTATTCCCGAGTATGATCCTA 240
 DB 956 AGATGCTCAGTGCCTTGTGGATGCTGAGCCCCCATCTATTCCCGAGTATGATCCTA 1015
 QY 241 CCAGACCCCTTCAGTGAAGCTTCGATGATGGGCTTACTGACCAACCTGGCAGACAGGAGC 300
 DB 1016 CCAGACCCCTTCAGTGAAGCTTCGATGATGGGCTTACTGACCAACCTGGCAGACAGGAGC 1075
 QY 301 TGGTTCAATGATCAAACTGGGCGAAGAGGGTGCCAG 336
 DB 1076 TGGTTCAATGATCAAACTGGGCGAAGAGGGTGCCAG 1111

RESULT 11
 AAD64590
 ID AAD64590 standard; DNA; 2028 BP.
 XX
 AC AAD64590;
 XX
 DT 12-FEB-2004 (first entry)
 DE Human ER3M-VP16-C-79 mutant fusion DNA.
 KW Tamoxifen activated system; TAS; 4-hydroxytamoxifen; OHT; gene therapy;
 KW oestrogen receptor; ER; fundamental research; transgenic; biomedical;
 KW disease model; breast cancer; transactivator; VP16; chimeric; human;
 KW mutant; gene; ds.
 XX
 OS Homo sapiens.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 CDS 1..2028
 /tag= a
 /product= "Human ER3M-VP16-C-79 mutant fusion protein"
 /note= "CDS does not include stop codon"
 FT
 FT misc_feature 1792..2028
 /partial
 /tag= b
 /note= "Human VP16-C-79 DNA"
 US2003199022-A1.
 23-OCT-2003.
 11-MAR-2002; 2002US-00095373.
 11-MAR-2002; 2002US-00095373.
 (UNII) UNIV ILLINOIS FOUND.
 Mao C, Shapiro DJ;
 WPI; 2003-852787/79.
 P-PSDB; ABW02643.
 XX
 PT New isolated polynucleotide useful for gene therapy applications (e.g. in
 PT treating breast cancer), in vivo and in vitro gene expression, in
 PT producing bioactive or toxic polypeptides, in research, or in producing
 PT transgenic animals.

```
XX PS Claim 22; SEQ ID NO 72; Opp; English.
XX CC
XX CC The invention relates to tamoxifen activated system (TAS) and 4-
XX CC hydroxytamoxifen (OHT) activated system for regulated production of
XX CC proteins in eukaryotic cells. TAS includes mutant oestrogen receptors
XX CC (ERs) and chimeras thereof. The invention is useful in gene therapy (e.g.
XX CC in treating breast cancer) or in vivo and in vitro gene expression, in
XX CC producing bioactive, toxic, recombinant polypeptides in mammalian cells,
XX CC in biomedical and fundamental research, or in producing transgenic
XX CC animals. The transgenic animals may be used as disease models, in
XX CC studying the function and/or activity of a polypeptide, or in identifying
XX CC and/or evaluating modulators of a polypeptide activity. The present
XX CC sequence is human ER3M-human transactivator VP16-C-79 mutant fusion DNA
XX CC used in TAS
XX SQ Sequence 2028 BP; 455 A; 588 C; 589 G; 396 T; 0 U; 0 Other;

Query Match 100.0%; Score 336; DB 10; Length 2028;
Best Local Similarity 100.0%; Pred. No. 1.3e-91;
Matches 336; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGATACGAAAGACCGAAGAGGAGGAGATGTTGAAACACAAAGCCGAGAGATGATG 60
Db 761 GGATACGAAAGACCGAAGAGGAGGAGATGTTGAAACACAAAGCCGAGAGATGATG 820
QY 61 GGGAGGCGAGGGGTGAAGTGGGTCTGCTGGACATGAGAGCTGCCAACCTTTGGCCAA 120
Db 821 GGGAGGCGAGGGGTGAAGTGGGTCTGCTGGACATGAGAGCTGCCAACCTTTGGCCAA 880
QY 121 GCCCGCTCATGATCAACCGCTCTAAGAGAACAGCGCTGGCTTGCCTGACGCGGAC 180
Db 881 GCCCGCTCATGATCAACCGCTCTAAGAGAACAGCGCTGGCTTGCCTGACGCGGAC 940
QY 181 AGATGGTCACTGCGCTTGTGGATGCTGAGCCCCCATCTACTTATTCGAGTATGATCCTA 240
Db 941 AGATGGTCACTGCGCTTGTGGATGCTGAGCCCCCATCTACTTATTCGAGTATGATCCTA 1000
QY 241 CCAGACCCCTTCAGTGAAGCTTCGATGATGGCTTACTGACCAACCTTGGCAGAGGAGC 300
Db 1001 CCAGACCCCTTCAGTGAAGCTTCGATGATGGCTTACTGACCAACCTTGGCAGAGGAGC 1060
QY 301 TGGTTTCATGATCAACTGGCGGAAGAGGGTGCCAG 336
Db 1061 TGGTTTCATGATCAACTGGCGGAAGAGGGTGCCAG 1096

RESULT 12
AAD64588
ID AAD64588 standard; DNA; 2031 BP.
XX AC
XX DT
XX DE 12-FEB-2004 (first entry)
XX DE Human ER3M-VP16-N-79 mutant fusion DNA.
XX DE
XX KW Tamoxifen activated system; TAS; 4-hydroxytamoxifen; OHT; gene therapy;
XX KW oestrogen receptor; ER; fundamental research; transgenic; biomedical;
XX KW disease model; breast cancer; transactivator; VP16; chimeric; human;
XX KW mutant; gene; ds.
XX OS Homo sapiens.
XX OS Synthetic.
XX PH
XX FT CDS
XX FT Location/Qualifiers
XX FT 1..2031
XX FT /*tag= a
XX FT /product= "Human ER3M-VP16-N-79 mutant fusion protein"
XX FT /note= "CDS does not include stop codon"
XX FT /partial
XX FT 4..240
XX FT misc_feature
XX FT /*tag= b

/Note= "Human VP16-N-79 DNA"
US2003199022-A1.
23-OCT-2003.
11-MAR-2002; 2002US-00095373.
11-MAR-2002; 2002US-00095373.
(UNII ) UNIV ILLINOIS FOUND.
Mao C, Shapiro DJ;
WPI: 2003-852787/79.
P-PSDB; ABW02641.
New isolated polynucleotide useful for gene therapy applications (e.g. in
treating breast cancer), in vivo and in vitro gene expression, in
producing bioactive or toxic polypeptides, in research, or in producing
transgenic animals.
Claim 22; SEQ ID NO 68; Opp; English.
The invention relates to tamoxifen activated system (TAS) and 4-
hydroxytamoxifen (OHT) activated system for regulated production of
proteins in eukaryotic cells. TAS includes mutant oestrogen receptors
(ERs) and chimeras thereof. The invention is useful in gene therapy (e.g.
in treating breast cancer) or in vivo and in vitro gene expression, in
producing bioactive, toxic, recombinant polypeptides in mammalian cells,
in biomedical and fundamental research, or in producing transgenic
animals. The transgenic animals may be used as disease models, in
studying the function and/or activity of a polypeptide, or in identifying
and/or evaluating modulators of a polypeptide activity. The present
sequence is human ER3M-human transactivator VP16-N-79 mutant fusion DNA
used in TAS
SQ Sequence 2031 BP; 456 A; 588 C; 589 G; 398 T; 0 U; 0 Other;

Query Match 100.0%; Score 336; DB 10; Length 2031;
Best Local Similarity 100.0%; Pred. No. 1.3e-91;
Matches 336; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGATACGAAAGACCGAAGAGGAGGAGATGTTGAAACACAAAGCCGAGAGATGATG 60
Db 1007 GGATACGAAAGACCGAAGAGGAGGAGATGTTGAAACACAAAGCCGAGAGATGATG 1066
QY 61 GGGAGGCGAGGGGTGAAGTGGGTCTGCTGGACATGAGAGCTGCCAACCTTTGGCCAA 120
Db 1067 GGGAGGCGAGGGGTGAAGTGGGTCTGCTGGACATGAGAGCTGCCAACCTTTGGCCAA 1126
QY 121 GCCCGCTCATGATCAACCGCTCTAAGAGAACAGCGCTGGCTTGCCTGACGCGGAC 180
Db 1127 GCCCGCTCATGATCAACCGCTCTAAGAGAACAGCGCTGGCTTGCCTGACGCGGAC 1186
QY 181 AGATGGTCACTGCGCTTGTGGATGCTGAGCCCCCATCTACTTATTCGAGTATGATCCTA 240
Db 1187 AGATGGTCACTGCGCTTGTGGATGCTGAGCCCCCATCTACTTATTCGAGTATGATCCTA 1246
QY 241 CCAGACCCCTTCAGTGAAGCTTCGATGATGGCTTACTGACCAACCTTGGCAGAGGAGC 300
Db 1247 CCAGACCCCTTCAGTGAAGCTTCGATGATGGCTTACTGACCAACCTTGGCAGAGGAGC 1306
QY 301 TGGTTTCATGATCAACTGGCGGAAGAGGGTGCCAG 336
Db 1307 TGGTTTCATGATCAACTGGCGGAAGAGGGTGCCAG 1342

RESULT 13
AAN70880
ID AAN70880 standard; cDNA; 2092 BP.
XX AC
XX AC AAN70880;
```

XX 25-MAR-2003 (revised)
DT 24-FEB-1991 (first entry)
XX
DE cDNA encoding human oestrogen receptor protein.
XX
KW Human oestrogen receptor protein; ss cDNA; expression system; assay;
KW steroid receptor protein.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 293..1280
FT /*tag= a
XX
XX W08705049-A.
XX
XX 27-AUG-1987.
XX
XX 18-FEB-1987; 87WO-US000341.
XX
XX 20-FEB-1986; 86US-00833829.
XX
XX (CALD) CALIFORNIA BIOTECHNOLOGY INC.
XX
XX Shine J;
XX
XX WPI; 1987-250213/35.
XX
XX P-PSDB; AAP70543.
XX
XX Expression system for vertebrate steroid receptor protein - comprising
XX DNA sequence encoding the protein linked to control sequences in
XX eucaryotic hosts.
XX
XX Disclosure; Fig 1-1 - 1-2; 26pp; English.
XX
XX The cDNA can be used in an expression system to express human oestrogen
XX receptor protein. The coding sequence is operably linked to control
XX sequences compatible with eukaryotic host cells. This method allows
XX expression under conditions which favour appropriate post-translational
XX processing. It produces large amts. of purified protein useful in the
XX design of agonist and antagonist cpds., for the study of the mechanism of
XX action of the steroid binding proteins in general, and for use in
XX diagnostic assays for the proteins or antibodies to them. These assays
XX are important in, eg the diagnosis of tumour sensitivities to steroid
XX metabolism. Suitable host cells are VERO, HeLa and CHO cells. (Updated on
XX 25-MAR-2003 to correct PA field.)
XX
SQ Sequence 2092 BP; 473 A; 607 C; 591 G; 421 T; 0 U; 0 Other;
Query Match 100.0%; Score 336; DB 1; Length 2092;
Best Local Similarity 100.0%; Pred. No. 1.3e-91;
Matches 336; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GGATACGAAAGACCGAAGAGGGGAGAGTGTGAAACACACAGCGCCAGAGAGATGATG 60
Db 1053 GGATACGAAAGACCGAAGAGGGGAGAGTGTGAAACACACAGCGCCAGAGAGATGATG 1112
Qy 61 GGGAGGCGCAGGGGTGAAGTGGGGTCTGCTGGAGACATGAGAGCTGCCAACCTTTGGCCAA 120
Db 1113 GGGAGGCGCAGGGGTGAAGTGGGGTCTGCTGGAGACATGAGAGCTGCCAACCTTTGGCCAA 1172
Qy 121 GCCCGCTCATGATCAACCGCTCTAAGAGAACAGCCTGGCCCTTCTCCCTGACGCCGAC 180
Db 1173 GCCCGCTCATGATCAACCGCTCTAAGAGAACAGCCTGGCCCTTCTCCCTGACGCCGAC 1232
Qy 181 AGATGGTCAGTGCCTTGTGGATGCTGAGCCCCCCTACTCTATTCCGAGTATGATCCTA 240
Db 1233 AGATGGTCAGTGCCTTGTGGATGCTGAGCCCCCCTACTCTATTCCGAGTATGATCCTA 1292
Qy 241 CCAGACCCCTTCAGTGAAGCTTCGATGATGGCTTACTGACCAACCTGGCAGACAGGAGC 300
Db 1293 CCAGACCCCTTCAGTGAAGCTTCGATGATGGCTTACTGACCAACCTGGCAGACAGGAGC 1352

Qy 301 TGGTTACATGATCAACTGGCGGAAGAGGGTCCAG 336
Db 1353 TGGTTACATGATCAACTGGCGGAAGAGGGTCCAG 1388

RESULT 14

AAC86920
ID AAC86920 standard; cDNA; 2092 BP.

XX
AC AAC86920;

XX
DT 02-APR-2001 (first entry)

XX
DE Nucleotide sequence of the human oestrogen receptor cDNA.

XX
KW Ribozyme; oestrogen-dependent tumour; cell proliferation; glucocorticoid;
KW DNA-binding domain; oestrogen receptor; cancer treatment; breast cancer;
XX ss.

OS Homo sapiens.

XX
PN W0200074485-A1.

XX
PD 14-DEC-2000.

XX
PF 02-JUN-2000; 2000WO-US015243.

XX
PR 04-JUN-1999; 99US-0137470P.

XX
PA (TEXA) UNIV TEXAS.

XX
PI Roy AK, Lavrovsky Y, Tyagi RK, Song CS, Chatterjee B;

XX
DR WPI; 2001-061633/07.

XX
PT Ribozyme having a high substrate specificity for an mRNA encoding a DNA-
PT binding domain of human estrogen receptor, useful for inhibiting estrogen
PT -dependent tumor cell proliferation, particularly breast cancer.

PS Disclosure; Page 8-9; 49pp; English.

XX The specification describes a ribozyme capable of inhibiting oestrogen-
XX dependent tumour cell proliferation and having a high substrate
XX specificity for an mRNA sequence encoding a DNA-binding domain of human
XX oestrogen receptor. The ribozyme is free of endonuclease activity for an
XX mRNA having a DNA binding domain of a glucocorticoid. The oestrogen
XX receptor site-specific ribozymes are useful for cancer treatment and
XX therapies, especially for inhibiting oestrogen-dependent tumour cell
XX proliferation, particularly breast cancer. The present sequence
XX represents the human oestrogen receptor cDNA

SQ Sequence 2092 BP; 473 A; 605 C; 593 G; 421 T; 0 U; 0 Other;

Query Match 100.0%; Score 336; DB 4; Length 2092;

Best Local Similarity 100.0%; Pred. No. 1.3e-91;

Matches 336; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGATACGAAAGACCGAAGAGGGGAGAGTGTGAAACACACAGCGCCAGAGAGATGATG 60

Db 1053 GGATACGAAAGACCGAAGAGGGGAGAGTGTGAAACACACAGCGCCAGAGAGATGATG 1112

Qy 61 GGGAGGCGCAGGGGTGAAGTGGGGTCTGCTGGAGACATGAGAGTGCCTTTGGCCAA 120

Db 1113 GGGAGGCGCAGGGGTGAAGTGGGGTCTGCTGGAGACATGAGAGTGCCTTTGGCCAA 1172

Qy 121 GCCCGCTCATGATCAACCGCTCTAAGAGAACAGCCTGGCCCTTCTCCCTGACGCCGAC 180

Db 1173 GCCCGCTCATGATCAACCGCTCTAAGAGAACAGCCTGGCCCTTCTCCCTGACGCCGAC 1232

Qy 181 AGATGGTCAGTGCCTTGTGGATGCTGAGCCCCCCTACTCTATTCCGAGTATGATCCTA 240

Db 1233 AGATGGTCAGTGCCTTGTGGATGCTGAGCCCCCCTACTCTATTCCGAGTATGATCCTA 1292

QY 241 CCAGACCCCTTCAGTGAAGCTTCGATGATGGGCTTACTGACCAACCTGGCGAGACAGGAGC 300
DB 1293 CCAGACCCCTTCAGTGAAGCTTCGATGATGGGCTTACTGACCAACCTGGCGAGACAGGAGC 1352
QY 301 TGGTTACATGATCAACTGGGCGAAGAGGGTGCAG 336
DB 1353 TGGTTACATGATCAACTGGGCGAAGAGGGTGCCAG 1388

RESULT 15
ABX89700
ID ABK89700 standard; DNA; 2092 BP.
XX AC ABK89700;
XX DT 05-NOV-2002 (first entry)
XX DE Oestrogen receptor alpha nucleic acid comprising A908G mutation, #6.
XX KW Oestrogen receptor alpha; breast cancer; pre-malignant lesion;
KW invasive breast cancer; A908G oestrogen receptor alpha transition;
KW oestrogen receptor alpha K303R substitution; human; gene; ds.
XX OS Homo sapiens.
XX PN WO200257283-A1.
XX PD 25-JUL-2002.
XX PF 16-JAN-2002; 2002WO-US004982.
XX PR 19-JAN-2001; 2001US-0262990P.
XX PR 09-JUL-2001; 2001US-0304018P.
XX PA (BAYU) BAYLOR COLLEGE MEDICINE.
XX PI Fuqua S, O'connell P, Allred DC, Hopp TA;
XX DR WPI; 2002-590711/63.
XX PT New isolated estrogen receptor alpha with A908G mutation or K303R
PT substitution, useful as diagnostic marker in breast tissue such as pre-
PT malignant lesions for the development of breast cancer, particularly
PT invasive breast cancer.

Claim 1; Page; 133pp; English.

CC The invention relates to an isolated oestrogen receptor alpha nucleic
CC acid sequence comprising an A908G mutation, or an amino acid sequence
CC comprising a K303R substitution. Also described are methods for detecting
CC susceptibility to development of breast cancer or invasive breast cancer
CC in an individual, for diagnosing breast cancer in an individual; and for
CC screening for a modulator of an oestrogen receptor alpha polypeptide
CC comprising a K303R substitution. The oestrogen receptor alpha is useful
CC as a diagnostic marker in breast tissue such as pre-malignant lesions
CC for the development of breast cancer, particularly invasive breast
CC cancer. The methods are useful for determining susceptibility to
CC development of breast cancer, for diagnosing, preventing or treating
CC breast cancer. Transgenic mice may be used for screening and identifying
CC agents that interact with the oestrogen receptor alpha, or affect breast
CC tissue health. The A908G oestrogen receptor alpha transition is
CC frequently present in pre-malignant lesions of the breast and can occur
CC in the adjacent normal-appearing breast epithelium. The present sequence
CC represents an oestrogen receptor alpha nucleic acid which may comprise
CC the A908G mutation. Note: The present sequence is not shown in the
CC specification but was obtained by the indexer from GenBank using the
CC accession number given in the specification

XX SQ Sequence 2092 BP; 473 A; 605 C; 593 G; 421 T; 0 U; 0 Other;

Query Match 100.0%; Score 336; DB 6; Length 2092;
Best Local Similarity 100.0%; Pred. No. 1.3e-91;

Matches 336; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GCATACGAAACACCCGAGAGGAGGAGATCTTGAACACAAAGCGCCAGAGAGATGATG 60
DB 1053 GGATACGAAACACCCGAGAGGAGGAGATCTTGAACACAAAGCGCCAGAGAGATGATG 1112
QY 61 GGGAGGGCAGGGGTGAAGTGGGGTCTGTGGAGACATGAGAGCTGCACACCTTTGGCCAA 120
DB 1113 GGGAGGGCAGGGGTGAAGTGGGGTCTGTGGAGACATGAGAGCTGCACACCTTTGGCCAA 1172
QY 121 GCCCGCTCATGATCAAAACGCTCTAAGAGAAACAGCCTTGTCCCTGACGGCCGACC 180
DB 1173 GCCCGCTCATGATCAAAACGCTCTAAGAGAAACAGCCTTGTCCCTGACGGCCGACC 1232
QY 181 AGATGGTCAGTGCCTTGTGGATGCTGAGCCCCCATACTCTATTCCGAGTATGATCCTA 240
DB 1233 AGATGGTCAGTGCCTTGTGGATGCTGAGCCCCCATACTCTATTCCGAGTATGATCCTA 1292
QY 241 CCAGACCCCTTCAGTGAAGCTTCGATGATGGGTTACTGACCAACCTGGCGAGACAGGAGC 300
DB 1293 CCAGACCCCTTCAGTGAAGCTTCGATGATGGGTTACTGACCAACCTGGCGAGACAGGAGC 1352
QY 301 TGGTTACATGATCAACTGGGCGAAGAGGGTGCCAG 336
DB 1353 TGGTTACATGATCAACTGGGCGAAGAGGGTGCCAG 1388

Search completed: November 3, 2004, 06:59:27
Job time : 689.273 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 3, 2004, 05:58:29 ; Search time 154.335 Seconds
(without alignments)
1547.447 Million cell updates/sec

Title: US-09-933-267A-1_COPY_306168_306503

Perfect score: 336

Sequence: 1 ggatacgaagaaagccgaaga.....ctggggaagagggtgcag 336

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 824507 seqs, 35539441 residues

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents NA.*

1: /cgn2_6/ptodata/1/ina/5A COMB.seq.*

2: /cgn2_6/ptodata/1/ina/5B COMB.seq.*

3: /cgn2_6/ptodata/1/ina/6A COMB.seq.*

4: /cgn2_6/ptodata/1/ina/6B COMB.seq.*

5: /cgn2_6/ptodata/1/ina/PCUTS COMB.seq.*

6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	336	100.0	2322	3	US-08-564-264-2
2	336	100.0	6450	3	US-09-041-886-34
3	336	100.0	6450	4	US-08-453-998-1
4	336	100.0	6450	4	US-10-027-983-3
5	333	99.1	8566	4	US-10-027-983-10
6	332.8	99.0	4963	1	US-08-076-726-16
7	332.8	99.0	4963	1	US-08-260-452-9
8	332.8	99.0	4963	2	US-08-481-970-9
9	332.8	99.0	4963	2	US-08-897-719-9
10	332.8	99.0	4963	3	US-09-163-269-9
11	332.8	99.0	4963	4	US-09-281-674-9
12	253.8	75.5	1956	3	US-08-693-940-2
13	253.8	75.5	1956	4	US-09-566-660-2
14	186.2	55.4	1218	4	US-10-027-983-98
15	70.8	21.1	2764	4	US-08-893-666A-1
16	67	19.9	1458	2	US-08-836-620A-6
17	59	17.6	1251	4	US-09-608-088-2
18	59	17.6	1251	4	US-09-711-288-2
19	59	17.6	1257	4	US-09-608-088-20
20	59	17.6	1257	4	US-09-711-288-20
21	59	17.6	1434	4	US-09-608-088-1
22	59	17.6	1434	4	US-09-711-288-1
23	59	17.6	1460	2	US-08-836-620A-4
24	59	17.6	1647	3	US-09-139-617-2
25	59	17.6	1647	4	US-09-561-741A-2
26	59	17.6	1647	4	US-09-558-795-2
27	59	17.6	1898	4	US-09-608-088-24

28	59	17.6	1898	4	US-09-711-288-24
29	56.6	16.8	2568	2	US-08-836-620A-1
30	47	14.0	1615	3	US-09-040-508-1
31	47	14.0	1615	3	US-09-500-654-1
c 32	38.6	11.5	505	4	US-09-621-976-15639
33	37.8	11.2	144	4	US-10-027-983-97
c 34	37.8	11.2	7218	1	US-08-232-463-14
35	36.2	10.8	2218	4	US-09-814-915A-4
36	36.2	10.8	2539	4	US-09-620-312D-454
37	34.6	10.3	4339	4	US-09-484-370B-164
38	33.2	9.9	3715	3	US-09-041-886-10
c 39	32.4	9.6	19056	3	US-09-272-032-8
c 40	32.4	9.6	19056	4	US-09-443-218-8
c 41	31.2	9.3	91	4	US-09-513-999C-15962
c 42	31	9.2	676	4	US-09-270-767-1688
c 43	31	9.2	676	4	US-09-270-767-16970
c 44	31	9.2	702	4	US-09-270-767-27
c 45	31	9.2	702	4	US-09-270-767-15309

ALIGNMENTS

RESULT 1

US-08-564-264-2

; Sequence 2, Application US/08564264

; Patent No. 6040430

; GENERAL INFORMATION:

; APPLICANT: STEWART, Francis

; TITLE OF INVENTION: REGULATION OF SITE-SPECIFIC

; TITLE OF INVENTION: RECOMBINATION BY SITE-SPECIFIC RECOMBINASE/NUCLEAR

; NUMBER OF SEQUENCES: 2

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Nikaido, Marmelstein, Murray & Oram

; STREET: 655 Fifteenth Street N.W. Suite 330

; CITY: Washington

; STATE: D.C.

; COUNTRY: U.S.A.

; ZIP: 20005-5701

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/564,264

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/EP94/02088

; FILING DATE: 28-JUN-1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: EP 93 110 298.2

; FILING DATE: 28-JUN-1993

; ATTORNEY/AGENT INFORMATION:

; NAME: Murray, Robert B.

; REGISTRATION NUMBER: 22,980

; REFERENCE/DOCKET NUMBER: P564-5019

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202)638-5000

; TELEFAX: (202)638-4810

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2322 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; FEATURE:

; NAME/KEY: misc_feature

; LOCATION: (1-1269)

; OTHER INFORMATION: /note= "FLP recombinase domain."

Search

```

; NAME/KEY: misc_feature
; LOCATION: (1270-1284)
; OTHER INFORMATION: /note= "Linker peptide."
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1285-2322)
; OTHER INFORMATION: /note= "Estrogen binding domain."
; US-08-564-264-2

Query Match 100.0%; Score 336; DB 3; Length 2322;
Best Local Similarity 100.0%; Pred. No. 2e-99;
Matches 336; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGTATCGAAAGACCGAGAGGAGGGAGAGATGTTGAAACACAGCGGCCAGAGAGATGATG 60
Db 1295 GGTATCGAAAGAACCGAAGAGAGGGAGAGATGTTGAAACACAAGCGGCCAGAGAGATGATG 1354
Qy 61 GGAGGGCCAGGGGTGAAGTGGGGTCTGCTCGAGACATGAGAGCTGCCAACCTTTGGCCAA 120
Db 1355 GGAGGGCCAGGGGTGAAGTGGGGTCTGCTCGAGACATGAGAGCTGCCAACCTTTGGCCAA 1414
Qy 121 GCCCGCTCATGATCAAAACGCTCTAAGAAGAACAGCCTGGCCTTTGTCCTTGACGCCGAC 180
Db 1415 GCCCGCTCATGATCAAAACGCTCTAAGAAGAACAGCCTGGCCTTTGTCCTTGACGCCGAC 1474
Qy 181 AGATGTCAGTGCCCTGTTGGATGCTGAGCCCCCATACTCTATTCCGAGTATGATCCTA 240
Db 1475 AGATGTCAGTGCCCTGTTGGATGCTGAGCCCCCATACTCTATTCCGAGTATGATCCTA 1534
Qy 241 CCAGACCCCTTCAGTGAAGCTTCGATGATGGCTTACTGACCAACCTGGCAGACAGGGAGC 300
Db 1535 CCAGACCCCTTCAGTGAAGCTTCGATGATGGCTTACTGACCAACCTGGCAGACAGGGAGC 1594
Qy 301 TGGTTCACATGATCAACTGGCGGCAAGAGGTGCCAG 336
Db 1595 TGGTTCACATGATCAACTGGCGGCAAGAGGTGCCAG 1630

```

```

RESULT 2
US-09-041-886-34
; Sequence 34, Application US/09041886
; Patent No. 6235872
; GENERAL INFORMATION:
; APPLICANT: Bredesen, Dale E.
; APPLICANT: Rabizadeh, Sharroz
; TITLE OF INVENTION: Proapoptotic Peptides, Dependence
; TITLE OF INVENTION: Polypeptides and Methods of Use
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/041,886
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 2626
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 34:
; SEQUENCE CHARACTERISTICS:

```

```

; LENGTH: 6450 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 361..2146
; US-09-041-886-34

Query Match      100.0%; Score 336; DB 3; Length 6450;
Best Local Similarity 100.0%; Pred. No. 3.le-99;
Matches 336; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  GGATACGAAAGACCGAAGAGGAGGAGAGATGTTGTAACACACAAGCGCCACAGAGATGATG 60
Db      1121 GGATACGAAAGACCGAAGAGGAGGAGAGATGTTGTAACACACAAGCGCCACAGAGATGATG 1180

QY      61  GGGAGGCGAGGGGTGAAGTGGGGTCTGCTGGAGACATGAGAGCTGCCAACCTTTGGCCAA 120
Db      1181 GGGAGGCGAGGGGTGAAGTGGGGTCTGCTGGAGACATGAGAGCTGCCAACCTTTGGCCAA 1240

QY      121 GCCCGCTCATGATCAAAACGCTTAAGAAGAACAGCCTGGCCTTGTCCTCGACGGCCGAGC 180
Db      1241 GCCCGCTCATGATCAAAACGCTTAAGAAGAACAGCCTGGCCTTGTCCTCGACGGCCGAGC 1300

QY      181 AGATGTCTAGTGCCTTGTGGATGCTGAGCCCCCATACTCTATTCCGAGTATGATCCTA 240
Db      1301 AGATGTCTAGTGCCTTGTGGATGCTGAGCCCCCATACTCTATTCCGAGTATGATCCTA 1360

QY      241 CCAGACCCCTTCAGTGAAGCTTCCATGATGGGCTTACTGCCAACCTTGGCAGACAGGGAGC 300
Db      1361 CCAGACCCCTTCAGTGAAGCTTCCATGATGGGCTTACTGCCAACCTTGGCAGACAGGGAGC 1420

QY      301 TGGTTCCATGATCAACTGGGCGAAGAGGGTGCCAG 336
Db      1421 TGGTTCCATGATCAACTGGGCGAAGAGGGTGCCAG 1456

```

```

RESULT 3
US-08-453-998-1
; Sequence 1, Application US/08453998
; Patent No. 644438
; GENERAL INFORMATION:
; APPLICANT: CHAMRON, PIERRE
; APPLICANT: METZGER, DANIEL
; APPLICANT: WHITE, JOHN
; TITLE OF INVENTION: METHOD FOR THE PREPARATION OF A PROTEIN
; TITLE OF INVENTION: BY YEASTS USING AN INDUCIBLE SYSTEM, VECTORS AND
; TITLE OF INVENTION: CORRESPONDING TRANSFORMED STRAINS
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN, DARBY & CUSHMAN
; STREET: 1100 NEW YORK AVE., N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/453,998
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/161,064
; FILING DATE: 03-DEC-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: CHAPIN, MARLANA K.
; REGISTRATION NUMBER: 35,843

```

Db	1181	GGGAGGGCAGGGGTGAAGTGGGCTCTGCTGGAGACATGAGAGCTGCCAACCTTTGGCCAA	1240
Qy	121	GCCCGCTCATGATCAAAACGCTCTTAAGAAACAGAGCTGGCCCTGTCCCTGACGGCCGACC	180
Db	1241	GCCCGCTCATGATCAAAACGCTCTTAAGAAACAGAGCTGGCCCTGTCCCTGACGGCCGACC	1300
Qy	181	AGATGGTCAAGTGCCTGTTGGATGCTGAGCCCCCATCTATTCGAGTATGATCCTA	240
Db	1301	AGATGGTCAAGTGCCTGTTGGATGCTGAGCCCCCATCTATTCGAGTATGATCCTA	1360
Qy	241	CCAGACCCCTTCAGTGAAGCTTCGATGATGGGCTTACTGACCAACCTGGCAGACAGGGAGC	300
Db	1361	CCAGACCCCTTCAGTGAAGCTTCGATGATGGGCTTACTGACCAACCTGGCAGACAGGGAGC	1420
Qy	301	TGGTTCACATGATCAACTGGCGGCAAGAGGTGGCAG	336
Db	1421	TGGTTCACATGATCAACTGGCGGCAAGAGGTGGCAG	1456

RESULT 5

US-10-027-983-10

US-10-027-983-10

Sequence 10, Application US/10027983

Patent No. 6617162

GENERAL INFORMATION:

APPLICANT: Kenneth W. Dobie

APPLICANT: Mark P. Roach

TITLE OF INVENTION: ANTISENSE MODULATION OF ESTROGEN RECEPTOR ALPHA EXPRESSION

FILE REFERENCE: RTS-0340

CURRENT APPLICATION NUMBER: US/10/027,983

CURRENT FILING DATE: 2001-12-18

NUMBER OF SEQ ID NOS: 98

SEQ ID NO 10

LENGTH: 8566

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: 5'UTR

LOCATION: (1)...(2302)

NAME/KEY: exon:exon junction

LOCATION: (1491)...(1492)

OTHER INFORMATION: exon 4:exon 5

NAME/KEY: start codon

LOCATION: (2302)...(2305)

NAME/KEY: 3'UTR

LOCATION: (2735)...(8566)

NAME/KEY: exon:exon junction

LOCATION: (3569)...(3570)

OTHER INFORMATION: exon 8:exon 9

NAME/KEY: exon:exon junction

LOCATION: (3708)...(3709)

OTHER INFORMATION: exon 9:exon 10

NAME/KEY: exon:exon junction

LOCATION: (3842)...(3843)

OTHER INFORMATION: exon 10:exon 11

NAME/KEY: exon:exon junction

LOCATION: (4026)...(4027)

OTHER INFORMATION: exon 11:exon 12

US-10-027-983-10

Query Match 99.1%; Score 333; DB 4; Length 8566;

Best Local Similarity 100.0%; Pred.No. 3.4e-98;

Matches 333; Conservative 0; Mismatches 0; Indels 0; Gaps 0

Qy	1	GGATACGAAAAAGACCGAGAGGAGGAGGAGATGTTGAACACACAAGCGCCAGAGAGATGATG	60
Db	3237	GGATACGAAAGACCGAGAGGAGGAGGAGATGTTGAACACACAAGCGCCAGAGAGATGATG	3299
Qy	61	GGGAGGGCAGGGGTGAAGTGGGGTCTGTGAGACATGAGAGCTGCCAACCTTTGGCCAA	120
Db	3297	GGGAGGGCAGGGGTGAAGTGGGGTCTGTGAGACATGAGAGCTGCCAACCTTTGGCCAA	3359
Qy	121	GCCCGCTCATGATCAAAACGCTCTTAAGAAACAGAGCTGGCCCTGTCCCTGACGGCCGACC	180

Db 1286 GGGAGGGCAGGGGTGAAGTGGGGTCTGCTGGAGACATGAGAGCTGCCAACCTTTGGCCAA 1345
QY 121 GCCCGCTCATGATCAAAAGCTCTAAGAGAACAGAGCTGGCCCTTGTCCCTGACGGCGGACC 180
Db 1346 GCCCGCTCATGATCAAAAGCTCTAAGAGAACAGAGCTGGCCCTTGTCCCTGACGGCGGACC 1405
QY 181 AGATGGTCAAGTCCCTTGTGGATGCTGAGCGCCGCCCATCTCTATTCCGAGTATGATCCTA 240
Db 1406 AGATGGTCAAGTCCCTTGTGGATGCTGAGCGCCGCCCATCTCTATTCCGAGTATGATCCTA 1465
QY 241 CCAGACCCCTTCAGTGAAGCTTCGATGATGGCTTACTGACCAACCTGGCAGACAGGGAGC 300
Db 1466 CCAGACCCCTTCAGTGAAGCTTCGATGATGGCTTACTGACCAACCTGGCAGACAGGGAGC 1525
QY 301 TGGTTCCATGATCAACTGGCGGAGAGGGGTGCCAG 336
Db 1526 TGGTTCCATGATCAACTGGCGGAGAGGGGTGCCAG 1561
RESULT 8
US-08-481-970-9
; Sequence 9, Application US/08481970
; Patent No. 5859310
; GENERAL INFORMATION:
; APPLICANT: Gossen, Manfred
; APPLICANT: Bujard, Hermann
; APPLICANT: Salfeld, Jochen
; APPLICANT: Voss, Jeffrey
; TITLE OF INVENTION: Animal Transgenic for a Tetracycline-Controlled
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lahive & Cockfield
; STREET: 60 State Street, Suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/481,970
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/260,452
; FILING DATE: 14-JUN-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/076,327
; FILING DATE: 14-JUN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Giulio A. DeConti, Jr.
; REGISTRATION NUMBER: 31,503
; REFERENCE/DOCKET NUMBER: BBI-013CP2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4963 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Human cytomegalovirus
; IMMEDIATE SOURCE:
; CLONE: pUHD BGR4
; US-08-481-970-9

Query Match 99.0%; Score 332.8; DB 2; Length 4963;
Best Local Similarity 99.4%; Pred. No. 3e-98;
Matches 334; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 GGATACGAAAGACCAAGAGAGAGGAGAGATGTTGAAACACAAAGCGCCAGAGAGATGATG 60
Db 1226 GGATACGAAAGACCAAGAGAGAGGAGAGATGTTGAAACACAAAGCGCCAGAGAGATGATG 1285
QY 61 GGGAGGGCAGGGGTGAAGTGGGGTCTGCTGGAGACATGAGAGCTGCCAACCTTTGGCCAA 120
Db 1286 GGGAGGGCAGGGGTGAAGTGGGGTCTGCTGGAGACATGAGAGCTGCCAACCTTTGGCCAA 1345
QY 121 GCCCGCTCATGATCAAAAGCTCTAAGAGAACAGAGCTGGCCCTTGTCCCTGACGGCGGACC 180
Db 1346 GCCCGCTCATGATCAAAAGCTCTAAGAGAACAGAGCTGGCCCTTGTCCCTGACGGCGGACC 1405
QY 181 AGATGGTCAAGTCCCTTGTGGATGCTGAGCGCCGCCCATCTCTATTCCGAGTATGATCCTA 240
Db 1406 AGATGGTCAAGTCCCTTGTGGATGCTGAGCGCCGCCCATCTCTATTCCGAGTATGATCCTA 1465
QY 241 CCAGACCCCTTCAGTGAAGCTTCGATGATGGCTTACTGACCAACCTGGCAGACAGGGAGC 300
Db 1466 CCAGACCCCTTCAGTGAAGCTTCGATGATGGCTTACTGACCAACCTGGCAGACAGGGAGC 1525
QY 301 TGGTTCCATGATCAACTGGCGGAGAGGGGTGCCAG 336
Db 1526 TGGTTCCATGATCAACTGGCGGAGAGGGGTGCCAG 1561
RESULT 9
US-08-897-719-9
; Sequence 9, Application US/08897719
; Patent No. 5922927
; GENERAL INFORMATION:
; APPLICANT: Gossen, Manfred
; APPLICANT: Bujard, Hermann
; APPLICANT: Salfeld, Jochen
; APPLICANT: Voss, Jeffrey
; TITLE OF INVENTION: Tight Control of Gene Expression in Eucaryotic
; TITLE OF INVENTION: Cells by Tetracycline-responsive Promoters
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lahive & Cockfield
; STREET: 60 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/897,719
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/260,452
; FILING DATE:
; APPLICATION NUMBER: 08/076,327
; FILING DATE: 14-JUN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Giulio A. DeConti, Jr.
; REGISTRATION NUMBER: 31,503
; REFERENCE/DOCKET NUMBER: BBI-013CP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4963 base pairs
; TYPE: nucleic acid

```
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Human cytomegalovirus
; IMMEDIATE SOURCE:
; CLONE: pUHD BGR4
US-08-897-719-9

Query Match          99.0%; Score 332.8; DB 2; Length 4963;
Best Local Similarity 99.4%; Pred. No. 3e-98; 2; Indels 0; Gaps 0;
Matches 334; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGATACGAAAGACCGAAGAGGAGGAGGAGATGTTGAAACACAAAGCGCCAGAGAGATGATG 60
Db 1226 GGATACGAAAGACCGAAGAGGAGGAGGAGATGTTGAAACACAAAGCGCCAGAGAGATG 1285

QY 61 GGGAGGCGAGGGGTGAAGTGGGCTCTGCTGGAGACATGAGAGCTGCCAACCTTTGGCCAA 120
Db 1286 GGGAGGCGAGGGGTGAAGTGGGCTCTGCTGGAGACATGAGAGCTGCCAACCTTTGGCCAA 1345

QY 121 GCCCGCTCATGATCAAAACGCTCTAAGAAGAAACAGCGCTGGCTTGCTCCCTGACGGCGGACC 180
Db 1346 GCCCGCTCATGATCAAAACGCTCTAAGAAGAAACAGCGCTGGCTTGCTCCCTGACGGCGGACC 1405

QY 181 AGATGTCATGTCCTTGTGGATGCTGAGCCCCCATCTCTATTCCGAGTATGATCCTA 240
Db 1406 AGATGTCATGTCCTTGTGGATGCTGAGCCCCCATCTCTATTCCGAGTATGATCCTA 1465

QY 241 CCAGACCCCTTCAGTGAAGCTTCGATGATGGCTTACTGACCAACCTGGCAGACAGGAGC 300
Db 1466 CCAGACCCCTTCAGTGAAGCTTCGATGATGGCTTACTGACCAACCTGGCAGACAGGAGC 1525

QY 301 TGGTTCACATGATCAACTGGGCGAAGAGGGTGCAG 336
Db 1526 TGGTTCACATGATCAACTGGGCGAAGAGGGTGCAG 1561

RESULT 10
US-09-163-269-9
; Sequence 9, Application US/09163269
; Patent No. 6252136
; GENERAL INFORMATION:
; APPLICANT: Gossen, Manfred
; APPLICANT: Bujard, Hermann
; APPLICANT: Salfeld, Jochen
; APPLICANT: Voss, Jeffrey
; TITLE OF INVENTION: Animal Transgenic for a Tetracycline-Controlled
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSSEE: Lahive & Cockfield
; STREET: 60 State Street, Suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/163,269
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/481,970
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/076,327
; FILING DATE: 14-JUN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Giulio A. DeConti, Jr.
```

```
; REGISTRATION NUMBER: 31,503
; REFERENCE/DOCKET NUMBER: BBI-013CP2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4963 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Human cytomegalovirus
; IMMEDIATE SOURCE:
; CLONE: pUHD BGR4
US-09-163-269-9

Query Match          99.0%; Score 332.8; DB 3; Length 4963;
Best Local Similarity 99.4%; Pred. No. 3e-98; 2; Indels 0; Gaps 0;
Matches 334; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGATACGAAAGACCGAAGAGGAGGAGGAGATGTTGAAACACAAAGCGCCAGAGAGATGATG 60
Db 1226 GGATACGAAAGACCGAAGAGGAGGAGGAGATGTTGAAACACAAAGCGCCAGAGAGATGATG 1285

QY 61 GGGAGGCGAGGGGTGAAGTGGGCTCTGCTGGAGACATGAGAGCTGCCAACCTTTGGCCAA 120
Db 1286 GGGAGGCGAGGGGTGAAGTGGGCTCTGCTGGAGACATGAGAGCTGCCAACCTTTGGCCAA 1345

QY 121 GCCCGCTCATGATCAAAACGCTCTAAGAAGAAACAGCGCTGGCTTGCTCCCTGACGGCGGACC 180
Db 1346 GCCCGCTCATGATCAAAACGCTCTAAGAAGAAACAGCGCTGGCTTGCTCCCTGACGGCGGACC 1405

QY 181 AGATGTCATGTCCTTGTGGATGCTGAGCCCCCATCTCTATTCCGAGTATGATCCTA 240
Db 1406 AGATGTCATGTCCTTGTGGATGCTGAGCCCCCATCTCTATTCCGAGTATGATCCTA 1465

QY 241 CCAGACCCCTTCAGTGAAGCTTCGATGATGGCTTACTGACCAACCTGGCAGACAGGAGC 300
Db 1466 CCAGACCCCTTCAGTGAAGCTTCGATGATGGCTTACTGACCAACCTGGCAGACAGGAGC 1525

QY 301 TGGTTCACATGATCAACTGGGCGAAGAGGGTGCAG 336
Db 1526 TGGTTCACATGATCAACTGGGCGAAGAGGGTGCAG 1561

RESULT 11
US-09-281-674-9
; Sequence 9, Application US/09281674
; Patent No. 6783756
; GENERAL INFORMATION:
; APPLICANT: Gossen, Manfred
; APPLICANT: Bujard, Hermann
; APPLICANT: Salfeld, Jochen
; APPLICANT: Voss, Jeffrey
; TITLE OF INVENTION: Methods for Regulating Gene Expression
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSSEE: Lahive & Cockfield
; STREET: 60 State Street, Suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/281,674
; FILING DATE: 30-Mar-1999
```



```
; PRIOR APPLICATION NUMBER: 08/693,940
; PRIOR FILING DATE: 1996-08-07
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1956
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1953)
; OTHER INFORMATION: Multi-chimeric transactivating factor
US-09-566-660-2

Query Match      75.5%; Score 253.8; DB 4; Length 1956;
Best Local Similarity 97.4%; Pred. No. 1.1e-72;
Matches 258; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 72 GGTGAAGTGGGCTCTCTGGAGACATGAGAGCTGCCAACCTTTGGCCAAAGCCGCTCATG 131
Db 1000 GGTGGGGATCCATCTCTGTGAGACATGAGAGCTGCCAACCTTTGGCCAAAGCCGCTCATG 1059

QY 132 ATCAAAACGCTCTAAGAAAGACAGCTGGCTTCTCCCTGACCGCCGACAGATGGTCAATG 191
Db 1060 ATCAAAACGCTCTAAGAAAGACAGCTGGCTTCTCCCTGACCGCCGACAGATGGTCAATG 1119

QY 192 GCCTTGTTCGATGCTGAGCCCCCCTACTTATTCGAGTATGATCTTACCAGACCTTTC 251
Db 1120 GCCTTGTTCGATGCTGAGCCCCCCTACTTATTCGAGTATGATCTTACCAGACCTTTC 1179

QY 252 AGTGAAGCTTCGATGATGGCTTACTGACCAACCTGGCAGACAGGAGTGTTCACATG 311
Db 1180 AGTGAAGCTTCGATGATGGCTTACTGACCAACCTGGCAGACAGGAGTGTTCACATG 1239

QY 312 ATCAACTGGGCGAAGAGGGTGCCAG 336
Db 1240 ATCAACTGGGCGAAGAGGGTGCCAG 1264

RESULT 14
US-10-027-983-98
; Sequence 98, Application US/10027983
; Patent No. 6617162
; GENERAL INFORMATION:
; APPLICANT: Kenneth W. Dobie
; TITLE OF INVENTION: ANTISENSE MODULATION OF ESTROGEN RECEPTOR ALPHA EXPRESSION
; FILE REFERENCE: RTS-0340
; CURRENT APPLICATION NUMBER: US/10/027,983
; PRIOR FILING DATE: 2001-12-18
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 98
; LENGTH: 1218
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
US-10-027-983-98

Query Match      55.4%; Score 186.2; DB 4; Length 1218;
Best Local Similarity 98.4%; Pred. No. 9e-51;
Matches 188; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 146 GAAGAACAGCGCTGGCCCTTGCCTTGACGGCCGACAGATGTCAGTCCCTTGTGGATGC 205
Db 336 GGACACACAGCGCTGGCCCTTGCCTTGACGGCCGACAGATGTCAGTCCCTTGTGGATGC 395

QY 206 TGAGCCCCCATACCTCTATTCGAGATGATCCCTTACCAGACCTTTCAGTGAAGCTTCAT 265
Db 396 TGAGCCCCCATACCTCTATTCGAGATGATCCCTTACCAGACCTTTCAGTGAAGCTTCAT 455

QY 266 GATGGGCTTACTGACCAACCTTGGCAGACAGGGAGCTGGTTTACATGATCAACTGGGCGAA 325
Db 456 GATGGGCTTACTGACCAACCTTGGCAGACAGGGAGCTGGTTTACATGATCAACTGGGCGAA 515
```

```
QY 326 GAGGTCGCCAG 336
Db 516 GAGGTCGCCAG 526
```

RESULT 15

```
US-09-893-666A-1
; Sequence 1, Application US/09893666A
; Patent No. 6759568
```

```
; GENERAL INFORMATION:
; APPLICANT: YAMASHITA, ICHIRO
; TITLE OF INVENTION: High estrogen-sensitive medaka fish
; FILE REFERENCE: 210217US-620-7249-0
; CURRENT APPLICATION NUMBER: US/09/893,666A
; PRIOR FILING DATE: 2002-02-12
; PRIOR APPLICATION NUMBER: JP 2000-247729
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 2764
; TYPE: DNA
; ORGANISM: Oryzias latipes
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (211)..(1935)
; OTHER INFORMATION:
US-09-893-666A-1
```

```
Query Match      21.1%; Score 70.8; DB 4; Length 2764;
Best Local Similarity 64.8%; Pred. No. 5e-13;
Matches 105; Conservative 0; Mismatches 57; Indels 0; Gaps 0;
```

```
QY 175 CCGACCAAGATGTCAGTGCCTTGTGGATGCTGAGCCCCCCTACTTATTCGAGATG 234
Db 1022 CTGAGCAGGTGCTGCTCCTTTCAGGGCGCGAGCCCCCGATCTCTGCTCGGTGAGA 1081

QY 235 ATCCTTACCAGACCTTTCAGTGAAGCTTCGATGATGGGCTTACTGACCAACCTTGGCAGACA 294
Db 1082 AGTTGAGCCGACCGTACACCGAGGTTCACCATGATGACCTGTCTACCAGCATGGCAGACA 1141

QY 295 GGGAGCTGTTTCAATGATCAACTGGGCGAAGAGGGTGCCAG 336
Db 1142 AGGAGCTGTCCATGATGCTGCTGGCCAAAGAGTCCCGAG 1183
```

```
Search completed: November 3, 2004, 09:46:54
Job time : 156.335 secs
```

GenCore version 5.1.6

Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 3, 2004, 06:51:04 ; Search time 611.713 Seconds
(without alignments)

```

US-09-933-26/A-1_COPI_306168_306303
Title:
Accession:
Length:
Perfect score: 336
Sequence:
1 ggaatagcaaaaaaaccaaga.....ctggcgaagaaggtgcag 336

```

Scoring table: IDENTITY NUC

scoring cable: IDENT11_A00
Gapop 10.0 ; Gapext 1.0

Searched: 3413475 seqs, 2563800928 residues

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Minimum	DB seq	length:	0
Maximum	DB seq	length:	200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA: *

1:	/cgn2_6/ptodata/2/pubnpna/US07_PUBCOMB.seq.*
2:	/cgn2_6/ptodata/2/pubnpna/PCT_NEW_PUB.seq.*
3:	/cgn2_6/ptodata/2/pubnpna/US06_NEW_PUB.seq.*
4:	/cgn2_6/ptodata/2/pubnpna/US06_PUBCOMB.seq.*
5:	/cgn2_6/ptodata/2/pubnpna/US07_NEW_PUB.seq.*
6:	/cgn2_6/ptodata/2/pubnpna/PCTUS_PUBCOMB.seq.*
7:	/cgn2_6/ptodata/2/pubnpna/US08_NEW_PUB.seq.*
8:	/cgn2_6/ptodata/2/pubnpna/US08_PUBCOMB.seq.*
9:	/cgn2_6/ptodata/2/pubnpna/US09A_PUBCOMB.seq.*
10:	/cgn2_6/ptodata/2/pubnpna/US09B_PUBCOMB.seq.*
11:	/cgn2_6/ptodata/2/pubnpna/US09C_PUBCOMB.seq.*
12:	/cgn2_6/ptodata/2/pubnpna/US09_NEW_PUB.seq.*
13:	/cgn2_6/ptodata/2/pubnpna/US10A_PUBCOMB.seq.*
14:	/cgn2_6/ptodata/2/pubnpna/US10B_PUBCOMB.seq.*
15:	/cgn2_6/ptodata/2/pubnpna/US10C_PUBCOMB.seq.*
16:	/cgn2_6/ptodata/2/pubnpna/US10D_PUBCOMB.seq.*
17:	/cgn2_6/ptodata/2/pubnpna/US10E_PUBCOMB.seq.*
18:	/cgn2_6/ptodata/2/pubnpna/US10_NEW_PUB.seq.*
19:	/cgn2_6/ptodata/2/pubnpna/US11_NEW_PUB.seq.*
20:	/cgn2_6/ptodata/2/pubnpna/US60_NEW_PUB.seq.*
21:	/cgn2_6/ptodata/2/pubnpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	Query	Length	DB	ID	Description
1	336	100.0	1237	14	US-10-052-092-3	Sequence 3, Appli	
2	336	100.0	1237	15	US-10-437-107-3	Sequence 3, Appli	
3	336	100.0	1785	9	US-10-095-373A-1	Sequence 1, Appli	
4	336	100.0	1788	9	US-09-853-033-1	Sequence 1, Appli	
5	336	100.0	2028	15	US-10-095-373A-72	Sequence 72, Appl	
6	336	100.0	2031	15	US-10-095-373A-68	Sequence 68, Appl	
7	336	100.0	2092	10	US-09-952-680A-9	Sequence 9, Appli	
8	336	100.0	2092	14	US-10-052-092-6	Sequence 6, Appli	
9	336	100.0	2092	15	US-10-137-899A-54	Sequence 54, Appl	
10	336	100.0	2092	15	US-10-437-107-6	Sequence 6, Appli	
11	336	100.0	2106	15	US-10-052-092-5	Sequence 5, Appli	
12	336	100.0	2106	15	US-10-437-107-5	Sequence 5, Appli	

13	336	100.0	2178	15	US-10-095-373A-70	Sequence 70, Appl
14	336	100.0	2181	15	US-10-095-373A-66	Sequence 66, Appl
15	336	100.0	6450	13	US-10-096-710-2	Sequence 2, Appl
16	336	100.0	6450	13	US-10-081-563-1	Sequence 1, Appl
17	336	100.0	6450	14	US-10-053-092-1	Sequence 1, Appl
18	336	100.0	6450	14	US-10-052-092-7	Sequence 7, Appl
19	336	100.0	6450	15	US-10-207-655-60	Sequence 60, Appl
20	336	100.0	6450	15	US-10-177-293-127	Sequence 127, App
21	336	100.0	6450	15	US-10-027-983-3	Sequence 3, Appl
22	336	100.0	6450	15	US-10-007-926A-422	Sequence 422, App
23	336	100.0	6450	15	US-10-437-107-1	Sequence 1, Appl
24	336	100.0	6450	15	US-10-437-107-7	Sequence 7, Appl
25	336	100.0	6450	15	US-10-392-274-1	Sequence 1, Appl
26	336	100.0	6450	15	US-10-448-753-3	Sequence 3, Appl
27	336	100.0	6450	15	US-10-172-118-435	Sequence 435, App
28	336	100.0	6450	15	US-10-388-360-303	Sequence 303, App
29	336	100.0	6450	16	US-10-342-887-335	Sequence 335, App
30	336	100.0	6610	14	US-10-053-092-20	Sequence 20, Appl
31	336	100.0	6610	15	US-10-437-107-20	Sequence 20, Appl
32	336	100.0	6823	15	US-10-422-934-10	Sequence 10, Appl
33	336	100.0	6956	15	US-10-422-934-9	Sequence 9, Appl
34	336	100.0	7038	15	US-10-422-934-3	Sequence 3, Appl
35	336	100.0	7979	18	US-10-473-433-9	Sequence 9, Appl
36	336	100.0	11365	17	US-10-433-311A-17	Sequence 17, Appl
37	336	100.0	4652337	9	US-09-933-2676A-1	Sequence 1, Appl
38	333	99.1	8566	15	US-10-027-983-10	Sequence 10, Appl
39	333	99.1	8566	15	US-10-448-753-10	Sequence 10, Appl
40	332.8	99.0	4963	9	US-09-281-674-9	Sequence 9, Appl
41	332.8	99.0	4963	9	US-09-777-317-9	Sequence 9, Appl
42	332.8	99.0	4963	9	US-09-892-227-9	Sequence 9, Appl
43	325	96.7	1317	15	US-10-157-899A-3	Sequence 3, Appl
44	325	96.7	1317	15	US-10-157-899A-11	Sequence 11, Appl
45	325	96.7	1317	15	US-10-157-899A-13	Sequence 13, Appl

ALIGNMENTS

```

RESULT 1
US-10-052-092-3
; Sequence 3, Application US/10052092
; Publication NO. US2003027778A1
; GENERAL INFORMATION:
; APPLICANT: Fucua, Suzanne
; APPLICANT: Alfred, D.
; APPLICANT: Hopp, Torsten A.
; APPLICANT: O'Connell, Peter
; TITLE OF INVENTION: Methods and Composition
; FILE REFERENCE: P02102US2
; CURRENT APPLICATION NUMBER: US/10/052,092,092
; CURRENT FILING DATE: 2002-01-18
; PRIOR APPLICATION NUMBER: US 60/262,990
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: US 60/304,018
; PRIOR FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 1237
; TYPE: DNA
; ORGANISM: Human
US-10-052-092-3

```

```

Query Match      100.0%; Score 336; DB 14; Length 1237;
Best Local Similarity 100.0%; Pred. No. 3.1e-104;
Matches 336; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1  GGATACGAAAGACCGAAGAGGGGAGAGTGTGTAACACAACGCGCCAGAGAGATGATG 60
Db      350 GGATACGAAAGACCGAAGAGGGGAGAGTGTGTAACACAACGCGCCAGAGAGATGATG 409

Qy      61  GGGAGGGCGAGGGGTGAAGTGGGTGCTCTGGAGACATGAGAGCTGCCAACCTTTGGCCAA 120

```

Db 410 GGGAGGCGAGGGGTGAAGTGGGGTCTGTGGAGACATGAGAGCTGCAACCTTTGGCCAA 469
QY 121 GCCCGCTCATGATCAAAACGCTCTAAGAAGAAACAGCCCTTGTCCCTTGACGCGCGACC 180
Db 470 GCCCGCTCATGATCAAAACGCTCTAAGAAGAAACAGCCCTTGTCCCTTGACGCGCGACC 529
QY 181 AGATGTCAGTGCCTTGTGGATGCTGAGCCCCCCTATCTCTATTCGAGTATGATCCTTA 240
Db 530 AGATGTCAGTGCCTTGTGGATGCTGAGCCCCCCTATCTCTATTCGAGTATGATCCTTA 589
QY 241 CCAGACCCCTTCAGTGAAGCTTCGATGATGGCTTACTGACCAACCTGGCAGACAGGAGC 300
Db 590 CCAGACCCCTTCAGTGAAGCTTCGATGATGGCTTACTGACCAACCTGGCAGACAGGAGC 649
QY 301 TGGTTCACATGATCAACTGGGCGAAGAGGGTGCCAG 336
Db 650 TGGTTCACATGATCAACTGGGCGAAGAGGGTGCCAG 685

RESULT 2

US-10-437-107-3
; Sequence 3, Application US/10437107
; Publication No. US20030186313A1
; GENERAL INFORMATION:
; APPLICANT: Fuqua, Suzanne
; APPLICANT: Allred, D.
; APPLICANT: Hopp, Torsten A.
; APPLICANT: O'Connell, Peter
; TITLE OF INVENTION: Methods and Composition in Breast Cancer Diagnosis and Therapeutic
; FILE REFERENCE: P02102US2
; CURRENT APPLICATION NUMBER: US/10/437,107
; CURRENT FILING DATE: 2003-05-13
; PRIOR APPLICATION NUMBER: US/10/052,092
; PRIOR FILING DATE: 2002-01-18
; PRIOR APPLICATION NUMBER: US 60/262,990
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: US 60/304,018
; PRIOR FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 1237
; TYPE: DNA
; ORGANISM: Human
US-10-437-107-3

Query Match 100.0%; Score 336; DB 15; Length 1237;
Best Local Similarity 100.0%; Pred. No. 3.1e-104; Indels 0; Gaps 0;
Matches 336; Conservative 0; Mismatches 0;

QY 1 GGATACGAAAGACCGAAGAGGAGGAGAGTGTGAAACACAAAGCGCCAGAGAGATGATG 60
Db 350 GGATACGAAAGACCGAAGAGGAGGAGAGTGTGAAACACAAAGCGCCAGAGAGATGATG 409
QY 61 GGGAGGCGAGGGTGAAGTGGGGTCTGTGGAGACATGAGAGCTGCCAACCTTTGGCCAA 120
Db 410 GGGAGGCGAGGGTGAAGTGGGGTCTGTGGAGACATGAGAGCTGCCAACCTTTGGCCAA 469
QY 121 GCCCGCTCATGATCAAAACGCTCTAAGAAGAAACAGCTTGTCCCTGACGCGCGACC 180
Db 470 GCCCGCTCATGATCAAAACGCTCTAAGAAGAAACAGCTTGTCCCTGACGCGCGACC 529
QY 181 AGATGTCAGTGCCTTGTGGATGCTGAGCCCCCCTATCTCTATTCGAGTATGATCCTTA 240
Db 530 AGATGTCAGTGCCTTGTGGATGCTGAGCCCCCCTATCTCTATTCGAGTATGATCCTTA 589
QY 241 CCAGACCCCTTCAGTGAAGCTTCGATGATGGCTTACTGACCAACCTGGCAGACAGGAGC 300
Db 590 CCAGACCCCTTCAGTGAAGCTTCGATGATGGCTTACTGACCAACCTGGCAGACAGGAGC 649
QY 301 TGGTTCACATGATCAACTGGGCGAAGAGGGTGCCAG 336
Db 650 TGGTTCACATGATCAACTGGGCGAAGAGGGTGCCAG 685

RESULT 3

US-10-095-373A-1
; Sequence 1, Application US/10095373A
; Publication No. US20030199022A1
; GENERAL INFORMATION:
; APPLICANT: Shapiro, David J.
; APPLICANT: Mao, Chengjian
; TITLE OF INVENTION: A TAMOXIFEN AND 4-HYDROXY TAMOXIFEN-ACTIVATED SYSTEM FOR REGULATING
; TITLE OF INVENTION: PRODUCTION OF PROTEINS IN EUKARYOTIC CELLS
; FILE REFERENCE: 10322.25
; CURRENT APPLICATION NUMBER: US/10/095,373A
; CURRENT FILING DATE: 2002-06-21
; NUMBER OF SEQ ID NOS: 83
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1785
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-095-373A-1

Query Match 100.0%; Score 336; DB 15; Length 1785;
Best Local Similarity 100.0%; Pred. No. 3.4e-104; Indels 0; Gaps 0;
Matches 336; Conservative 0; Mismatches 0;

QY 1 GGATACGAAAGACCGAAGAGGAGGAGAGTGTGAAACACAAAGCGCCAGAGAGATGATG 60
Db 761 GGATACGAAAGACCGAAGAGGAGGAGAGTGTGAAACACAAAGCGCCAGAGAGATGATG 820
QY 61 GGGAGGCGAGGGTGAAGTGGGGTCTGTGGAGACATGAGAGCTGCCAACCTTTGGCCAA 120
Db 821 GGGAGGCGAGGGTGAAGTGGGGTCTGTGGAGACATGAGAGCTGCCAACCTTTGGCCAA 880
QY 121 GCCCGCTCATGATCAAAACGCTCTAAGAAGAAACAGCTTGTCCCTGACGCGCGACC 180
Db 881 GCCCGCTCATGATCAAAACGCTCTAAGAAGAAACAGCTTGTCCCTGACGCGCGACC 940
QY 181 AGATGTCAGTGCCTTGTGGATGCTGAGCCCCCCTATCTCTATTCGAGTATGATCCTTA 240
Db 941 AGATGTCAGTGCCTTGTGGATGCTGAGCCCCCCTATCTCTATTCGAGTATGATCCTTA 1000
QY 241 CCAGACCCCTTCAGTGAAGCTTCGATGATGGGGTGTACTGACCAACCTGGCAGACAGGAGC 300
Db 1001 CCAGACCCCTTCAGTGAAGCTTCGATGATGGGGTGTACTGACCAACCTGGCAGACAGGAGC 1060
QY 301 TGGTTCACATGATCAACTGGGCGAAGAGGGTGCCAG 336
Db 1061 TGGTTCACATGATCAACTGGGCGAAGAGGGTGCCAG 1096

RESULT 4

US-09-853-033-1
; Sequence 1, Application US/09853033
; Patent No. US20020100069A1
; GENERAL INFORMATION:
; APPLICANT: CHAMBER, PIERRE
; APPLICANT: METZGER, DANIEL
; TITLE OF INVENTION: TRANSGENIC MOUSE FOR TARGETED RECOMBINATION
; TITLE OF INVENTION: MEDIATED BY MODIFIED CRE-ER
; FILE REFERENCE: 065691/0222
; CURRENT APPLICATION NUMBER: US/09/853,033
; CURRENT FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: FR 00/12570
; PRIOR FILING DATE: 2000-10-03
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1788
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS


```

; PRIOR APPLICATION NUMBER: 60/232,454
; PRIOR FILING DATE: 2000-09-13
; NUMBER OF SEQ ID NOS: 75
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 9
; LENGTH: 2092
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-952-680A-9

```

	Query Match	100.0%;	Score 336;	DB 10;	Length 2092;
	Best Local Similarity	100.0%;	Pred. No. 3.6e-104;		
	Matches 336;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	GGATACAAAGACCGAAGAGGGGAGAGATGTTGAACACACAGCGCCAGAGAGATGATG	60		
Db	1053	GGATACAAAGACCGAAGAGGGGAGAGATGTTGAACACACAGCGCCAGAGAGATGATG	1112		
Qy	61	GGGAGGCGCAGGGGTGAAGTGGGGTCTGCTGGAGACATGAGAGCTGCCAAACCTTTGGCCAA	120		
Db	1113	GGGAGGCGCAGGGGTGAAGTGGGGTCTGCTGGAGACATGAGAGCTGCCAAACCTTTGGCCAA	1172		
Qy	121	GCOCGCTCATGATCAAAACGCTCTAAGAAGAACAGCCTTGGCCTTGTCTCGAGCGCCGACC	180		
Db	1173	GCOCGCTCATGATCAAAACGCTCTAAGAAGAACAGCCTTGGCCTTGTCTCGAGCGCCGACC	1232		
Qy	181	AGATGGTCACTGCTTGTGGATGCTGAGCGCCGCCATACCTCTATTCCGAGTATGATCCTA	240		
Db	1233	AGATGGTCACTGCTTGTGGATGCTGAGCGCCGCCATACCTCTATTCCGAGTATGATCCTA	1292		
Qy	241	CCAGACCCCTTCAGTGAAGCTTCGATGATGGCTTACTTGACCAACCTGGCAGACAGGGAGC	300		
Db	1293	CCAGACCCCTTCAGTGAAGCTTCGATGATGGCTTACTTGACCAACCTGGCAGACAGGGAGC	1352		
Qy	301	TGGTTCATGATCAACTGGCGCAGAGGGTGCAG	336		
Db	1353	TGGTTCATGATCAACTGGCGCAGAGGGTGCAG	1388		

RESULT 8
 US-10-052-092-6
 ; Sequence 6, Application US/10052092
 ; Publication No. US200302778A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Fucua, Suzanne
 ; APPLICANT: Allred, D.
 ; APPLICANT: Hopp, Torsten A.
 ; APPLICANT: O'Connell, Peter
 ; TITLE OF INVENTION: Methods and Composition in Breast Cancer Diagnosis and Therapeutic
 ; FILE REFERENCE: P02102US2
 ; CURRENT APPLICATION NUMBER: US/10/052,092
 ; CURRENT FILING DATE: 2002-01-18
 ; PRIOR APPLICATION NUMBER: US 60/262,990
 ; PRIOR FILING DATE: 2001-01-19
 ; PRIOR APPLICATION NUMBER: US 60/304,018
 ; PRIOR FILING DATE: 2001-07-09
 ; NUMBER OF SEQ ID NOS: 49
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 6
 ; LENGTH: 2092
 ; TYPE: DNA
 ; ORGANISM: Human
 ; US-10-052-092-6

Query Match	100.0%;	Score 336;	DB 14;	Length 2092;
Best Local Similarity	100.0%;	Pred. NO. 3.6e-104;		
Matches 336;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	GGATACAAAGACCGAAGAGGGAGGAGAAATGTTGAAACACAAAGCGCCAGAGATGATG	60	
Db	1053	GGATACAAAGACCGAAGAGGGAGGAGAAATGTTGAAACACAAAGCGCCAGAGATGATG	1112	
Dv	61	GGGAGGCGAGGGGTGAAGTGGGGTCTGCTGGAGACATGAGAGCTGCCAACCTTTGGCCAA	120	

Db	1113	GGAGGCGCAGGGTGAAGTGGGTCTGTGGAGCATGAGAGCTGCAACCTTTGGCCAA	1172
Qy	121	GCCCGCTCATGATCAAAACGCTCPAGAAGAACAGCTTGGCCTTGTTCCTGAACGGCCGAC	180
Db	1173	GCCCGCTCATGATCAAAACGCTCTAAGAAGAACAGCTTGGCCTTGTCTGCTGACGGCCGAC	1232
Qy	181	AGATGTCAGTGCCTTGTGTGGATGCTGAGCCCCCATCTCTATTCGGAGTAGATGCTTA	240
Db	1233	AGATGTCAGTGCCTTGTGTGGATGCTGAGCCCCCATCTCTATTCGGAGTAGATGCTTA	1292
Qy	241	CCAGACCTTCAGTGAAGCTTCGATGATGGCTTACTGACCAACCTGGCAGACAGGAGC	300
Db	1293	CCAGACCTTCAGTGAAGCTTCGATGATGGCTTACTGACCAACCTGGCAGACAGGAGC	1352
Qy	301	TGGTTTCACATGATCAACTGGGCGAAGAGGTGCCAG	336
Db	1353	TGGTTTCACATGATCAACTGGGCGAAGAGGTGCCAG	1388

RESULT 9

```

US-10-157-899A-54
Sequence 54, Application US/10157899A
Publication No. US20030143559A1
GENERAL INFORMATION:
APPLICANT: Bracken, Kathryn Rene
APPLICANT: de los Angeles, Joseph Ernest
APPLICANT: Huang, Ying
APPLICANT: Kadan, Michael Joseph
APPLICANT: Keander, Gary Michael
APPLICANT: Zerby, Dennis
TITLE OF INVENTION: NOVEL ESTROGEN RECEPTOR
TITLE OF INVENTION: LIGANDS AND PHARMACEUT
FILE REFERENCE: 4-32019A
CURRENT APPLICATION NUMBER: US/10/157,899A
CURRENT FILING DATE: 2002-09-12
PRIOR APPLICATION NUMBER: US 60/294,839
PRIOR FILING DATE: 2001-05-31
NUMBER OF SEQ ID NOS: 55
SOFTWARE: PatentIn version 3.1
SEQ ID NO 54
LENGTH: 2092
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (293)..(2080)
OTHER INFORMATION:
FEATURE:
NAME/KEY: misc feature
LOCATION: (293)..(832)
OTHER INFORMATION: A/B domain
FEATURE:
NAME/KEY: misc feature
LOCATION: (833)..(1081)
OTHER INFORMATION: C domain
FEATURE:
NAME/KEY: misc feature
LOCATION: (1082)..(1198)
OTHER INFORMATION: D domain
FEATURE:
NAME/KEY: misc feature
LOCATION: (1199)..(1954)
OTHER INFORMATION: E domain
FEATURE:
NAME/KEY: misc feature
LOCATION: (1955)..(2077)
OTHER INFORMATION: F domain
US-10-157-899A-54

```

```
Query Match      100.0%; Score 336; DB 15; Length 2092;
Best Local Similarity 100.0%; Pred. No. 3.6e-104;
Matches 336; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```


QY 1 GGATACGAAAGACCGAGAGGGGAGGAATGTTGAAACACACGCGCCGAGAGATGATG 60
Db 1053 GGTATACGAAAGACCGAGAGGGGAGGAATGTTGAAACACACGCGCCGAGAGATGATG 1112
QY 61 GGGAGGCGAGGGGTGAAGTGGGGTCTCTGGAGACATGAGAGCTGCCAACTTTGGCCAA 120
Db 1113 GGGAGGCGAGGGGTGAAGTGGGGTCTCTGGAGACATGAGAGCTGCCAACTTTGGCCAA 1172
QY 121 GCCCGTCTATGATCAAAACGCTCTAAGAAAGAACAGCCTTGGCTTGTCCCTGACGGCCGACC 180
Db 1173 GCCCGTCTATGATCAAAACGCTCTAAGAAAGAACAGCCTTGGCTTGTCCCTGACGGCCGACC 1232
QY 181 AGATGGTCAGTGCCTTGTGGATGCTCAGCCCCCATCTACTTATCCGAGTATGATCCTA 240
Db 1233 AGATGGTCAGTGCCTTGTGGATGCTCAGCCCCCATCTACTTATCCGAGTATGATCCTA 1292
QY 241 CCAGACCTTCAGTGAAGCTTCGATGATGGGCTTACTGACCAACTGGCAGACAGGGAGC 300
Db 1293 CCAGACCTTCAGTGAAGCTTCGATGATGGGCTTACTGACCAACTGGCAGACAGGGAGC 1352
QY 301 TGGTTCAATGATCAACTGGGCGAAGAGGGGTGCCAG 336
Db 1353 TGGTTCAATGATCAACTGGGCGAAGAGGGGTGCCAG 1388

RESULT 10

US-10-437-107-6
; Sequence 6, Application US/10437107
; Publication No. US20030186313A1
; GENERAL INFORMATION:
; APPLICANT: Fuqua, Suzanne
; APPLICANT: Allred, D.
; APPLICANT: Hopp, Torsten A.
; APPLICANT: O'Connell, Peter
; TITLE OF INVENTION: Methods and Composition in Breast Cancer Diagnosis and Therapeutic
; FILE REFERENCE: P02102US2
; CURRENT APPLICATION NUMBER: US/10/437,107
; CURRENT FILING DATE: 2003-05-13
; PRIOR APPLICATION NUMBER: US/10/052,092
; PRIOR FILING DATE: 2002-01-18
; PRIOR APPLICATION NUMBER: US 60/262,990
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: US 60/304,018
; PRIOR FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 2092
; TYPE: DNA
; ORGANISM: Human
US-10-437-107-6

Query Match 100.0%; Score 336; DB 15; Length 2092;
Best Local Similarity 100.0%; Pred. No. 3.6e-104;
Matches 336; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GGATACGAAAGACCGAGAGGGGAGGAATGTTGAAACACACGCGCCGAGAGATGATG 60
Db 1053 GGTATACGAAAGACCGAGAGGGGAGGAATGTTGAAACACACGCGCCGAGAGATGATG 1112
QY 61 GGGAGGCGAGGGGTGAAGTGGGGTCTCTGGAGACATGAGAGCTGCCAACTTTGGCCAA 120
Db 1113 GGGAGGCGAGGGGTGAAGTGGGGTCTCTGGAGACATGAGAGCTGCCAACTTTGGCCAA 1172
QY 121 GCCCGTCTATGATCAAAACGCTCTAAGAAAGAACAGCCTTGGCTTGTCCCTGACGGCCGACC 180
Db 1173 GCCCGTCTATGATCAAAACGCTCTAAGAAAGAACAGCCTTGGCTTGTCCCTGACGGCCGACC 1232
QY 181 AGATGGTCAGTGCCTTGTGGATGCTGAGCCCCCATCTACTTATCCGAGTATGATCCTA 240
Db 1233 AGATGGTCAGTGCCTTGTGGATGCTGAGCCCCCATCTACTTATCCGAGTATGATCCTA 1292

QY 241 CCAGACCTTCAGTGAAGCTTCGATGATGGGCTTACTGACCAACTGGCAGACAGGGAGC 300
Db 1293 CCAGACCTTCAGTGAAGCTTCGATGATGGGCTTACTGACCAACTGGCAGACAGGGAGC 1352
QY 301 TGGTTCAATGATCAACTGGGCGAAGAGGGGTGCCAG 336
Db 1353 TGGTTCAATGATCAACTGGGCGAAGAGGGGTGCCAG 1388

RESULT 11

US-10-052-092-5
; Sequence 5, Application US/10052092
; Publication No. US20030027778A1
; GENERAL INFORMATION:
; APPLICANT: Fuqua, Suzanne
; APPLICANT: Allred, D.
; APPLICANT: Hopp, Torsten A.
; APPLICANT: O'Connell, Peter
; TITLE OF INVENTION: Methods and Composition in Breast Cancer Diagnosis and Therapeutic
; FILE REFERENCE: P02102US2
; CURRENT APPLICATION NUMBER: US/10/052,092
; CURRENT FILING DATE: 2002-01-18
; PRIOR APPLICATION NUMBER: US 60/262,990
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: US 60/304,018
; PRIOR FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 2106
; TYPE: DNA
; ORGANISM: Human
US-10-052-092-5

Query Match 100.0%; Score 336; DB 14; Length 2106;

Best Local Similarity 100.0%; Pred. No. 3.6e-104;
Matches 336; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGATACGAAAGACCGAGAGGGGAGGAATGTTGAAACACACGCGCCGAGAGATGATG 60
Db 761 GGATACGAAAGACCGAGAGGGGAGGAATGTTGAAACACACGCGCCGAGAGATGATG 820
QY 61 GGGAGGCGAGGGGTGAAGTGGGGTCTCTGGAGACATGAGAGCTGCCAACTTTGGCCAA 120
Db 821 GGGAGGCGAGGGGTGAAGTGGGGTCTCTGGAGACATGAGAGCTGCCAACTTTGGCCAA 880
QY 121 GCCCGTCTATGATCAAAACGCTCTAAGAAAGAACAGCCTTGGCTTGTCCCTGACGGCCGACC 180
Db 881 GCCCGTCTATGATCAAAACGCTCTAAGAAAGAACAGCCTTGGCTTGTCCCTGACGGCCGACC 940
QY 181 AGATGGTCAGTGCCTTGTGGATGCTGAGCCCCCATCTACTTATCCGAGTATGATCCTA 240
Db 941 AGATGGTCAGTGCCTTGTGGATGCTGAGCCCCCATCTACTTATCCGAGTATGATCCTA 1000
QY 241 CCAGACCTTCAGTGAAGCTTCGATGATGGGCTTACTGACCAACTGGCAGACAGGGAGC 300
Db 1001 CCAGACCTTCAGTGAAGCTTCGATGATGGGCTTACTGACCAACTGGCAGACAGGGAGC 1060
QY 301 TGGTTCAATGATCAACTGGGCGAAGAGGGGTGCCAG 336
Db 1061 TGGTTCAATGATCAACTGGGCGAAGAGGGGTGCCAG 1096

RESULT 12

US-10-437-107-5
; Sequence 5, Application US/10437107
; Publication No. US20030186313A1
; GENERAL INFORMATION:
; APPLICANT: Fuqua, Suzanne
; APPLICANT: Allred, D.
; APPLICANT: Hopp, Torsten A.
; APPLICANT: O'Connell, Peter
; TITLE OF INVENTION: Methods and Composition in Breast Cancer Diagnosis and Therapeutic

FILE REFERENCE: P02102US2
CURRENT APPLICATION NUMBER: US/10/437,107
CURRENT FILING DATE: 2003-05-13
PRIOR APPLICATION NUMBER: US/10/052,092
PRIOR FILING DATE: 2002-01-18
PRIOR APPLICATION NUMBER: US 60/262,990
PRIOR FILING DATE: 2001-01-19
PRIOR APPLICATION NUMBER: US 60/304,018
PRIOR FILING DATE: 2001-07-09
NUMBER OF SEQ ID NOS: 49
SOFTWARE: PatentIn version 3.1
SEQ ID NO 5
LENGTH: 2106
TYPE: DNA
ORGANISM: Human
US-10-437-107-5

Query Match 100.0%; Score 336; DB 15; Length 2106;
Best Local Similarity 100.0%; Pred. No. 3.6e-104;
Matches 336; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGATACGAAAGACCGAAGAGGAGGAGAGATGTTGAAACACAAAGCCGAGAGATGATG 60
Db 761 GGATACGAAAGACCGAAGAGGAGGAGAGATGTTGAAACACAAAGCCGAGAGATGATG 820
QY 61 GGGAGGCGAGGGGTGAAGTGGGCTCTGCTGGAGACATGAGAGCTGCCAACCTTTGGCCAA 120
Db 821 GGGAGGCGAGGGGTGAAGTGGGCTCTGCTGGAGACATGAGAGCTGCCAACCTTTGGCCAA 880
QY 121 GCCCGCTCATGATCAAAACCGCTCTAAGAGAACAGCGCTGGCCTTGTCCCTGACGGCGACC 180
Db 881 GCCCGCTCATGATCAAAACCGCTCTAAGAGAACAGCGCTGGCCTTGTCCCTGACGGCGACC 940
QY 181 AGATGTCAGTGCCTTGTGGATGCTGAGCCCCCATCTCTATTCCGAGTATGATCCTA 240
Db 941 AGATGTCAGTGCCTTGTGGATGCTGAGCCCCCATCTCTATTCCGAGTATGATCCTA 1000
QY 241 CCAGACCCCTTCAGTGAAGCTTCGATGATGGGCTTACTGACCAACCTGGCAGAGGGAGC 300
Db 1001 CCAGACCCCTTCAGTGAAGCTTCGATGATGGGCTTACTGACCAACCTGGCAGAGGGAGC 1060
QY 301 TGGTTACATGATCAACTGGCGAAGAGGGTGCCAG 336
Db 1061 TGGTTACATGATCAACTGGCGAAGAGGGTGCCAG 1096

RESULT 13

US-10-095-373A-70
Sequence 70, Application US/10095373A
Publication No. US20030199022A1

GENERAL INFORMATION:
APPLICANT: Shapiro, David J.
TITLE OF INVENTION: A TAMOXIFEN AND 4-HYDROXY TAMOXIFEN-ACTIVATED SYSTEM FOR REGULATING
TITLE OF INVENTION: PRODUCTION OF PROTEINS IN EUKARYOTIC CELLS
FILE REFERENCE: 10322.25
CURRENT APPLICATION NUMBER: US/10/095,373A
CURRENT FILING DATE: 2002-06-21
NUMBER OF SEQ ID NOS: 83
SOFTWARE: PatentIn version 3.1
SEQ ID NO 70

LENGTH: 2178

TYPE: DNA
ORGANISM: Artificial sequence
FEATURE:
OTHER INFORMATION: construct
US-10-095-373A-70

Query Match 100.0%; Score 336; DB 15; Length 2178;
Best Local Similarity 100.0%; Pred. No. 3.6e-104;
Matches 336; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGATACGAAAGACCGAAGAGGAGGAGAGATGTTGAAACACAAAGCCGAGAGATGATG 60

Db 761 GGATACGAAAGACCGAAGAGGAGGAGAGATGTTGAAACACAAAGCCGAGAGATGATG 820
QY 61 GGGAGGCGAGGGGTGAAGTGGGCTCTGCTGGAGACATGAGAGCTGCCAACCTTTGGCCAA 120
Db 821 GGGAGGCGAGGGGTGAAGTGGGCTCTGCTGGAGACATGAGAGCTGCCAACCTTTGGCCAA 880
QY 121 GCCCGCTCATGATCAAAACCGCTCTAAGAGAACAGCGCTGGCCTTGTCCCTGACGGCGACC 180
Db 881 GCCCGCTCATGATCAAAACCGCTCTAAGAGAACAGCGCTGGCCTTGTCCCTGACGGCGACC 940
QY 181 AGATGTCAGTGCCTTGTGGATGCTGAGCCCCCATCTCTATTCCGAGTATGATCCTA 240
Db 941 AGATGTCAGTGCCTTGTGGATGCTGAGCCCCCATCTCTATTCCGAGTATGATCCTA 1000
QY 241 CCAGACCCCTTCAGTGAAGCTTCGATGATGGGCTTACTGACCAACCTGGCAGAGGGAGC 300
Db 1001 CCAGACCCCTTCAGTGAAGCTTCGATGATGGGCTTACTGACCAACCTGGCAGAGGGAGC 1060
QY 301 TGGTTACATGATCAACTGGCGAAGAGGGTGCCAG 336
Db 1061 TGGTTACATGATCAACTGGCGAAGAGGGTGCCAG 1096

RESULT 14

US-10-095-373A-66
Sequence 66, Application US/10095373A
Publication No. US20030199022A1

GENERAL INFORMATION:
APPLICANT: Shapiro, David J.
TITLE OF INVENTION: A TAMOXIFEN AND 4-HYDROXY TAMOXIFEN-ACTIVATED SYSTEM FOR REGULATING
TITLE OF INVENTION: PRODUCTION OF PROTEINS IN EUKARYOTIC CELLS
FILE REFERENCE: 10322.25
CURRENT APPLICATION NUMBER: US/10/095,373A
CURRENT FILING DATE: 2002-06-21
NUMBER OF SEQ ID NOS: 83
SOFTWARE: PatentIn version 3.1
SEQ ID NO 66
LENGTH: 2181
TYPE: DNA
ORGANISM: Artificial sequence
FEATURE:
OTHER INFORMATION: construct
US-10-095-373A-66

Query Match 100.0%; Score 336; DB 15; Length 2181;
Best Local Similarity 100.0%; Pred. No. 3.6e-104;
Matches 336; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGATACGAAAGACCGAAGAGGAGGAGAGATGTTGAAACACAAAGCCGAGAGATGATG 60
Db 1157 GGATACGAAAGACCGAAGAGGAGGAGAGATGTTGAAACACAAAGCCGAGAGATGATG 1216
QY 61 GGGAGGCGAGGGGTGAAGTGGGCTCTGCTGGAGACATGAGAGCTGCCAACCTTTGGCCAA 120
Db 1217 GGGAGGCGAGGGGTGAAGTGGGCTCTGCTGGAGACATGAGAGCTGCCAACCTTTGGCCAA 1276
QY 121 GCCCGCTCATGATCAAAACCGCTCTAAGAGAACAGCGCTGGCCTTGTCCCTGACGGCGACC 180
Db 1277 GCCCGCTCATGATCAAAACCGCTCTAAGAGAACAGCGCTGGCCTTGTCCCTGACGGCGACC 1336
QY 181 AGATGTCAGTGCCTTGTGGATGCTGAGCCCCCATCTCTATTCCGAGTATGATCCTA 240
Db 1337 AGATGTCAGTGCCTTGTGGATGCTGAGCCCCCATCTCTATTCCGAGTATGATCCTA 1396
QY 241 CCAGACCCCTTCAGTGAAGCTTCGATGATGGGCTTACTGACCAACCTGGCAGAGGGAGC 300
Db 1397 CCAGACCCCTTCAGTGAAGCTTCGATGATGGGCTTACTGACCAACCTGGCAGAGGGAGC 1456
QY 301 TGGTTACATGATCAACTGGCGAAGAGGGTGCCAG 336
Db 1457 TGGTTACATGATCAACTGGCGAAGAGGGTGCCAG 1492

```
RESULT 15
US-10-096-710-2
; Sequence 2, Application US/10096710
; Publication No. US20020164581A1
; GENERAL INFORMATION:
; APPLICANT: Chang, Chawnsang
; TITLE OF INVENTION: METHODS FOR IDENTIFYING COMPOUNDS THAT
; TITLE OF INVENTION: DO NOT ATTENUATE THE PROTECTIVE EFFECTS OF ESTROGEN
; FILE REFERENCE: 21108.001002
; CURRENT APPLICATION NUMBER: US/10/096,710
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/275,229
; PRIOR FILING DATE: 2001-03-12
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 6450
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:/note =
; OTHER INFORMATION: synthetic construct
US-10-096-710-2

Query Match      100.0%; Score 336; DB 13; Length 6450;
Best Local Similarity 100.0%; Pred. No. 5e-104;
Matches 336; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCATACGAAAGACCGAGAGGAGAGATGTTGAAACACAAAGCGCCAGAGAGATGATG 60
Db 1121 GCATACGAAAGACCGAGAGGAGAGATGTTGAAACACAAAGCGCCAGAGAGATG 1180

QY 61 GGGAGGGCAGGGGTGAAGTGGGTCTCTGGAGACATGAGAGCTGCCAACCTTTGGCCAA 120
Db 1181 GGGAGGGCAGGGGTGAAGTGGGTCTCTGGAGACATGAGAGCTGCCAACCTTTGGCCAA 1240

QY 121 GCCCGCTCATGATCAAAAGCTCTAAGAGAACAGAGCTGGCCTTGCCCTGACGGCGGACC 180
Db 1241 GCCCGCTCATGATCAAAAGCTCTAAGAGAACAGAGCTGGCCTTGCCCTGACGGCGGACC 1300

QY 181 AGATGTCAGTCCCTTGTGGATGCTGAGCCCCCATCTCTATTCCGAGTATGATCCTA 240
Db 1301 AGATGTCAGTCCCTTGTGGATGCTGAGCCCCCATCTCTATTCCGAGTATGATCCTA 1360

QY 241 CCAGACCCCTTCAGTGAAGCTTCGATGATGGGCTTACTGACCAACCTGGCAGACAGGAGC 300
Db 1361 CCAGACCCCTTCAGTGAAGCTTCGATGATGGGCTTACTGACCAACCTGGCAGACAGGAGC 1420

QY 301 TGGTTCACATGATCAACTGGGCGAGAGGGTGCCAG 336
Db 1421 TGGTTCACATGATCAACTGGGCGAGAGGGTGCCAG 1456
```

Search completed: November 3, 2004, 10:01:04
Job time : 612.713 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 3, 2004, 04:57:13 ; Search time 5800.42 Seconds
(without alignments)
2110.838 Million cell updates/sec

Title: US-09-933-267A-1_COPY_306168_306503
Perfect score: 336
Sequence: 1 ggatacgaagaccgaaga.....ctggcgagagggtgccag 336

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

- 1: gb_est1:*
- 2: gb_est2:*
- 3: gb_hic:*
- 4: gb_est3:*
- 5: gb_est4:*
- 6: gb_est5:*
- 7: gb_est6:*
- 8: gb_gss1:*
- 9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	336	100.0	611	5	EX489897	EX489897 DKFZp686N0562
2	334.4	99.5	653	1	AI073549	AI073549 ov45e08.x
3	284.8	84.8	705	5	BP160291	BP160291 BP160291
4	257.6	76.7	616	4	EG084311	EG084311 H3099A08-
5	257.6	76.7	2109	3	AK039311	AK039311 Mus muscu
6	257.6	76.7	2591	3	AK077236	AK077236 Mus muscu
7	257.6	76.7	2735	3	AK054182	AK054182 Mus muscu
8	257.6	76.7	4248	3	AK041525	AK041525 Mus muscu
9	257.6	76.7	4315	3	AK087638	AK087638 Mus muscu
10	257.6	76.7	4321	3	AK036627	AK036627 Mus muscu
11	245.6	73.1	593	1	AA023625	AA023625 mh80b06.r
12	222.2	66.1	365	1	AA092184	AA092184 114609.se
13	202.8	60.4	627	1	BB200714	BB200714 BB200714
14	170.8	50.8	576	1	AU041270	AU041270 AU041270
15	118.4	35.2	318	2	BB097271	BB097271 BB097271
16	91.8	27.3	648	2	BF660250	BF660250 maa95912.
17	80.4	23.9	459	8	BL157672	BL157672 CH230-433
18	68.4	20.4	1203	9	CL025390	CL025390 CH216-21J
19	66.8	19.9	1277	9	CL025388	CL025388 CH216-21J
20	66.6	19.8	642	2	BF058527	BF058527 7k3ih08.x
21	65.4	19.5	2712	3	AK054290	AK054290 Mus muscu
22	65.4	19.5	3144	3	AK054301	AK054301 Mus muscu
23	65.4	19.5	3363	3	AK054413	AK054413 Mus muscu
24	64.4	19.2	1160	6	CD508522	CD508522 CDA91-C11

C	25	63.2	18.8	906	9	CNS01WT6	AL170835 Tetraodon
	26	61.8	18.4	641	2	AW976636	AW976636 EST388745
	27	60.4	18.0	735	8	BZ834605	BZ834605 CH240_209
	28	59.6	17.7	765	9	CNS03FXW	AL242285 Tetraodon
	29	58.6	17.4	642	2	BF733032	BF733032 nael18a05.
	30	58	17.3	992	9	CNS038DM	AL233483 Tetraodon
	31	55.8	16.6	955	9	CNS04QF6	AL302487 Tetraodon
	32	53	15.8	422	6	CA383629	CA383629 663832 NC
	33	52	15.5	275	7	CNS03ZWK	CN980443 48400_126
	34	48.6	14.5	971	9	CNS03ZWK	AL268157 Tetraodon
	35	47	14.0	303	4	BM723962	BM723962 UI-E-B01-
	36	47	14.0	696	4	BM723962	BM723962 UI-E-B01-
	37	47	14.0	1013	7	BM723962	BM723962 UI-E-B01-
	38	47	14.0	1377	9	AY412347	CN647617 ILLUMIGEN
	39	47	14.0	1377	9	AY412347	AY412347 Homo sapi
	40	46.8	13.9	714	9	CNS01VNY	AY412348 Pan trogl
	41	46.4	13.8	979	9	CNS04DMC	AL169351 Tetraodon
	42	45.4	13.5	312	4	EG211665	AL285933 Tetraodon
	43	45.4	13.5	548	7	CO627140	RG211665 RST31233
	44	44.4	13.2	366	4	EG194599	CO627140 D39-283d1
	45	44	13.1	1273	4	BI079710	EG194599 RST13761
							BI079710 602876202

ALIGNMENTS

RESULT 1
EX489897
LOCUS
DEFINITION
EX489897 611 bp mRNA linear EST 04-SEP-2003
DKFZp686N0562 r1 686 (synonym: hlcc3) Homo sapiens cDNA clone
DKFZp686N0562 5', mRNA sequence.
ACCESSION
EX489897
VERSION
EX489897.1
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
REFERENCE
1 (bases 1 to 611)
AUTHORS
Wambutt,R., Heubner,D., Mewes,H.W., Weil,B., Amid,C., Osanger,A., Fobo,G., Han,M. and Wiemann,S.
TITLE
EST (Wambutt R., Heubner D., Mewes H.W., Weil B., Amid C., et al.)
JOURNAL
Unpublished (2003)
COMMENT
Contact: MIPS
MIPS
Ingolstaedter Landstr.1, D-85764 Neuberberg, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by AGOWA (Berlin/Germany) within the cDNA sequencing consortium of the German Genome Project.
No sl sequence available.
This clone (DKFZp686N0562) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

FEATURES

source
Location/Qualifiers
1..611
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="DKFZp686N0562"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="686 (synonym: hlcc3)"
/note="Vector: pTriplex2; Site_1: SfiIA; Site_2: SfiIB;
cDNA-collection"

ORIGIN

Query Match 100.0%; Score 336; DB 5; Length 611;
Best Local Similarity 100.0%; Pred. No. 4.8e-84;
Matches 336; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 GGATACGAAGACCGAGGAGGAGATGTTGAAACACACAGCCGAGATGATG 60

Db 5 GGATACGAAAGACCGAGAGGGGAGATGTTGAAACACACAGCCGACGAGATGATG 64
QY 61 GGGAGGGGAGGGGTGAAGTGGGTCTGCTGGAGACATGAGAGCTGCCAACCTTTGGCCAA 120
Db 65 GGGAGGGGAGGGGTGAAGTGGGTCTGCTGGAGACATGAGAGCTGCCAACCTTTGGCCAA 124
QY 121 GCCCGCTCATGATCAACCGCTCTAAGAAGAACAGAGCTGGCTTGTCCCTGACGGCCGACC 180
Db 125 GCCCGCTCATGATCAACCGCTCTAAGAAGAACAGAGCTGGCTTGTCCCTGACGGCCGACC 184
QY 181 AGATGTCAGTGGCTTGTGGATGCTGAGCCCCCATCTCTATTCCGAGTATGATCCTA 240
Db 185 AGATGTCAGTGGCTTGTGGATGCTGAGCCCCCATCTCTATTCCGAGTATGATCCTA 244
QY 241 CCAGACCCCTTCAGTGAAGCTTCGATGATGGCTTACTGACCAACCTGGGACAGAGGAGC 300
Db 245 CCAGACCCCTTCAGTGAAGCTTCGATGATGGCTTACTGACCAACCTGGGACAGAGGAGC 304
QY 301 TGGTTCACATGATCAACTGGGCGAAGAGGGTGCCAG 336
Db 305 TGGTTCACATGATCAACTGGGCGAAGAGGGTGCCAG 340

RESULT 2
AI073549/c AI073549 653 bp mRNA linear EST 06-AUG-1998
LOCUS ov45e08.x1 Soares testis NHT Homo sapiens cDNA clone IMAGE:1640294
DEFINITION 3', similar to gb:M12674 ESTROGEN RECEPTOR (HUMAN);, mRNA sequence.
ACCESSION AI073549
VERSION AI073549.1 GI:3400193
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 653)
REFERENCE NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
cDNA library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CCAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 483.
Location/Qualifiers
1..653
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1640294"
/sex="male"
/lab_host="DH10B"
/clone_lib="Soares testis NHT"
/note="Vector: pT730-Pac (Pharmacia) with a modified
polylinker; Site1: Not I; Site 2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech
Laboratories, Inc., and primed with a Not I - oligo(dT)
primer [5']
TGTTACCAATCTGAAGTGGAGCGGCCGCCCAATTTTTTTTTTTT 3']
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization to Cot5, and was
constructed by Bento Soares and M. Fatima Bonaldo."

Query Match 99.5%; Score 334.4; DB 1; Length 653;
Best Local Similarity 99.7%; Pred. No. 1.4e-83;
Matches 335; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGATACGAAAGACCGAGAGGGGAGATGTTGAAACACACAGCCGACGAGATGATG 60
Db 652 GGATACGAAAGACCGAGAGGGGAGATGTTGAAACACACAGCCGACGAGATGATG 593
QY 61 GGGAGGGGAGGGGTGAAGTGGGTCTGCTGGAGACATGAGAGCTGCCAACCTTTGGCCAA 120
Db 592 GGGAGGGGAGGGGTGAAGTGGGTCTGCTGGAGACATGAGAGCTGCCAACCTTTGGCCAA 533
QY 121 GCCCGCTCATGATCAACCGCTCTAAGAAGAACAGAGCTGGCTTGTCCCTGACGGCCGACC 180
Db 532 GCCCGCTCATGATCAACCGCTCTAAGAAGAACAGAGCTGGCTTGTCCCTGACGGCCGACC 473
QY 181 AGATGTCAGTGGCTTGTGGATGCTGAGCCCCCATCTCTATTCCGAGTATGATCCTA 240
Db 472 AGATGTCAGTGGCTTGTGGATGCTGAGCCCCCATCTCTATTCCGAGTATGATCCTA 413
QY 241 CCAGACCCCTTCAGTGAAGCTTCGATGATGGCTTACTGACCAACCTGGGACAGAGGAGC 300
Db 412 CCAGACCCCTTCAGTGAAGCTTCGATGATGGCTTACTGACCAACCTGGGACAGAGGAGC 353
QY 301 TGGTTCACATGATCAACTGGGCGAAGAGGGTGCCAG 336
Db 352 TGGTTCACATGATCAACTGGGCGAAGAGGGTGCCAG 317

RESULT 3
BP160291 705 bp mRNA linear EST 30-DEC-2003
LOCUS BP160291 full-length enriched swine cDNA library, adult thymus Sus
DEFINITION BP160291 scrofa cDNA clone THY010029A10 5', mRNA sequence.
ACCESSION BP160291
VERSION BP160291.1 GI:40409764
KEYWORDS EST.
SOURCE Sus scrofa (pig)
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
1 (bases 1 to 705)
REFERENCE Uenishi,H., Eguchi,T., Suzuki,K., Sawazaki,T., Toki,D., Shinkai,H.,
Okumura,N., Hamasima,N. and Awata,T.
PEDE (Pig EST Data Explorer): construction of a database for ESTs
derived from porcine full-length cDNA libraries
Nucleic Acids Res. 32 (1), D484-D488 (2004)
Contact: Hirohide Uenishi
Animal Genome Laboratory, Genome Research Department
National Institute of Agrobiological Sciences
2 Ikenodai, Tsukuba, Ibaraki 305-8602, Japan
Tel: +81-29-838-8627
Fax: +81-29-838-8627
Email: huenishi@affrc.go.jp
EST project with full-length enriched cDNA libraries carried out in
Animal Genome Research Program (Japan) by National Institute of
Agrobiological Sciences and STAFF-Institute
Single pass sequencing of clones derived from oligo-capped cDNA
library
Vector sequences were eliminated by RepeatMasker version 2002/07/13
and crossmatch version 0.990319
Low quality bases were trimmed based on the quality values.
Location/Qualifiers
1..705
/organism="Sus scrofa"
/mol_type="mRNA"
/db_xref="taxon:9823"
/clone="THY010029A10"
/tissue type="thymus"
/dev stage="adult"
/clone_lib="full-length enriched swine cDNA library, adult
thymus"

ORIGIN

Query Match 84.8%; Score 284.8; DB 5; Length 705;
 Best Local Similarity 90.5%; Pred. No. 1.5e-69;
 Matches 304; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 1 GGATACGAAAGACCGAAGAGGAGGAGGAGATGTTGAAACACAAAGCCGACGAGAGATGATG 60
 Db |||||
 359 GGATACGAAAGACCGAAGAGGAGGAGGAGATGTTGAAACACAAAGCCGACGAGAGATGATG 418
 QY 61 GGAGAGCGGAGGGGTGAAGTGGGGTCTGCTGGAGACATGAGAGTGCACACCTTTGGCCAA 120
 Db |||||
 419 GGAGAGCGGAGGATGAAGCGTGGCCCTGGAGACATGAGATGCGCAACCTTTGGCCAA 478
 QY 121 GCGCGCTCATGATCAACAGCTCTTAAGAAGAACAGCCTGGCGTGTCTGCTGACGGCCGACC 180
 Db |||||
 479 GCGCTCTCTTGATTAACACACTAAGAAGAACAGCCTGGCGTGTCTGCTGACGGCCGACC 538
 QY 181 AGATGTCAGTGCCTTTGTTGGATGCTGAGCCGCCCATCTCTATTCGAGATGATCCTTA 240
 Db |||||
 539 AGATGATCAGTGCCTTTGTTGGAGGCTGAGCCGCCCATATCTATTCGAGATGATCCTTA 598
 QY 241 CCAGACCCCTTCAGTGAAGCTTCGATGATGCGTCTACTGACCACTGGCAGACAGGAGC 300
 Db |||||
 599 CCAGACCCCTTCAGTGAAGCTTCGATGATGCGTCTACTGACCACTGGCAGACAGGAGC 658
 QY 301 TGGTTCACATGATCAACTGGGCGAAGAGGGTGCAG 336
 Db |||||
 659 TGGCACACATGATCACTGGGCGAAGAGGGTGCAG 694

RESULT 4
 BG084311
 LOCUS
 DEFINITION H3099A08-5 NIA Mouse 15K cDNA Clone Set Mus musculus cDNA clone
 H3099A08 5', mRNA sequence.
 BG084311
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 Tanaka.T.S., Jaradat.S.A., Lim.M.K., Kargul.G.J., Wang.X.,
 Grahovac.M.J., Pantano.S., Sano.Y., Piao.Y., Nagaraja.R., Doi.H.,
 Wood.W.H. III, Becker.K.G. and Ko.M.S.H.
 Genome-wide expression profiling of mid-gestation placenta and
 embryo using a 15,000 mouse developmental cDNA microarray
 Proc. Natl. Acad. Sci. U.S.A. 97 (16), 9127-9132 (2000)
 20381348
 10922068
 On Jan 26, 2001 this sequence version replaced gi.12566875.
 Other ESTs: H3099A08-3
 Contact: George J. Kargul
 Laboratory of Genetics
 National Institute on Aging/National Institutes of Health
 333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
 Email: cdna@lgsun.grc.nia.nih.gov
 This clone set has been freely distributed to the community. Please
 visit <http://lgsun.grc.nia.nih.gov/cdna/15k.html> for details.
 Plate: H3099 row: A column: 08
 Seq primer: -21M13 Reverse
 High quality sequence stop: 616
 POLYA=No.

FEATURES
 Source
 1. .616
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57Bl/6J"
 /db_xref="niaST:H3099A08-5"
 /db_xref="taxon:10090"
 /clone="H3099A08"

/sex="Clones arrayed from a variety of cDNA libraries"
 /dev_stage="Clones arrayed from a variety of cDNA libraries"
 /lab_host="DH10B"
 /clone_lib="NIA Mouse 15K cDNA Clone Set"
 /notes="Vector: pSPORT1; Site 1: Sali; Site 2: NotI; This clone is among a rearrayed set of 15,247 clones from 11 embryo cDNA libraries (including preimplantation stage embryos from unfertilized egg to blastocyst, embryonic part of E7.5 embryos, extraembryonic part of E7.5 embryos, and E12.5 female mesonephros/gonad) and one newborn ovary cDNA library. Average insert size 1.5 kb. All source libraries are cloned unidirectionally with Oligo(dT)-Not primers. References include: (1) Genome-wide expression profiling of mid-gestation placenta and embryo using a 15,000 mouse developmental cDNA microarray, 2000, Proc. Natl. Acad. Sci. U.S.A. 97: 9127-9132; (2) Large-scale cDNA analysis reveals phased gene expression patterns during preimplantation mouse development, 2000, Development, 127: 1737-1749; (3) Genome-wide mapping of unselected transcripts from extraembryonic tissue of 7.5-day mouse embryos reveals enrichment in the t-complex and under-representation on the X chromosome, 1998, Hum Mol Genet 7: 1967-1978."

ORIGIN

Query Match 76.7%; Score 257.6; DB 4; Length 616;
 Best Local Similarity 85.4%; Pred. No. 7.3e-62;
 Matches 287; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 1 GGATACGAAAGACCGAAGAGGAGGAGATGTTGAAACACAAAGCCGACGAGATGATG 60
 Db |||||
 177 GCATACGAAAGACCGCAGGAGGAGGAGATGTTGAAACACAAAGCGTCAGAGATGATG 236
 QY 61 GGGAGGGCAGGGGTGAAGTGGGGTCTGCTGGAGACATGAGAGCTGCCAACCTTTGGCCAA 120
 Db |||||
 237 TGGAGGGCCGAAATGAAATGGTCTTCAGAGAGACATGAGGGTCCCAACCTTTGGCCAA 296
 QY 121 GCGCGCTCATGATCAACAGCTCTTAAGAAGAACAGCCTGGCGTGTCTGCTGACGGCCGACC 180
 Db |||||
 287 GCGCTCTCTTGATTAAGCACACTAAGAAGAAATAGCCCTGCTGCTTGCATGAGCTGACC 356
 QY 181 AGATGTCAGTGCCTTTGTTGGATGCTGAGCCGCCCATCTCTATTCGAGATGATCCTTA 240
 Db |||||
 357 AGATGTCAGTGCCTTTGTTGGATGCTGAGCCGCCCATGATCTATTCGATGATGATCCTT 416
 QY 241 CCAGACCCCTTCAGTGAAGCTTCGATGATGAGGGCTTACTGACCAACCTGGCAGACAGGAGC 300
 Db |||||
 417 CTAGACCCCTTCAGTGAAGCCCTCAATGATGGGCTTATTGACCAACCTAGCAGATAGGAGC 476
 QY 301 TGGTTCACATGATCAACTGGGCGAAGAGGGTGCAG 336
 Db |||||
 477 TGGTTCATGATCACTGGGCGAAGAGAGTGCAG 512

RESULT 5
 AK039911
 LOCUS
 DEFINITION Mus musculus 0 day neonate thymus cDNA, RIKEN full-length enriched library, clone:A430027L01 product:estrogen receptor 1 (alpha), full insert sequence.
 AK039911
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 Carninci,P. and Hayashizaki,Y.
 High-efficiency full-length cDNA cloning
 Meth. Enzymol. 303, 19-44 (1999)
 99279253

```

10349636
2
REFERENCE
AUTHORS
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Sasaki,N., Carninci,P.,
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL
Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE
20499374
PUBMED
11042159
REFERENCE
AUTHORS
3
Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
Konno,H., Akiyama,J., Nishi,K., Kitsuunai,T., Tashiro,H., Itoh,M.,
Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsunura,S., Kawai,J.,
Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
JOURNAL
Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE
20530913
PUBMED
11076861
REFERENCE
AUTHORS
4
The RIKEN Genome Exploration Research Group Phase II Team and the
PANTOM Consortium.
TITLE
Functional annotation of a full-length mouse cDNA collection
JOURNAL
Nature 409, 685-690 (2001)
REFERENCE
AUTHORS
5
The PANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
TITLE
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
JOURNAL
Nature 420, 563-573 (2002)
REFERENCE
AUTHORS
6
(bases 1 to 2109)
Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
Katon,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
Koya,S., Kurihara,C., Matsuyama,F., Miyazaki,A., Murata,M.,
Nakamura,M., Nishi,K., Nomura,K., Numata,K., Ohno,M., Ohsato,N.,
Okazaki,Y., Saichou,K., Saichou,H., Sakai,C., Sakai,K., Sakazume,N.,
Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
Muramatsu,M. and Hayashizaki,Y.
TITLE
Direct Submission
JOURNAL
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.jp,
URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222,
Fax:81-45-503-9216)
COMMENT
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL:http://genome.gsc.riken.jp/
URL:http://fantom.gsc.riken.jp/
FEATURES
source
Location/Qualifiers
1..2109
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:A430027L01"
/db_xref="taxon:10090"
/clone="A430027L01"
/tissue_type="thymus"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="0 day neonate"
misc_feature
1..2109
/note="estrogen receptor 1 (alpha) (MGI:1352467,
GB|NM_007956, evidence: BLASTN, 99%, match=1960)."
ORIGIN
Query Match
Best Local Similarity 76.7%; Score 257.6; DB 3; Length 2109;
Matches 287; Conservative 0; Mismatches 49; Indels 0; Gaps 0;
QY 1 GGATACGAAAGACGAGAGGAGGAGAGATGTGAAACACACAGCGCCAGAGAGATGATG 60
DB 1458 GATACGAAAGACGAGAGGAGGAGAGATGTGAAAGCACAAGCGTCAGAGAGATGACT 1517
QY 61 GGGAGGCGCAGGGGTGAAGTGGGGTCTGCTGGAGACATGAGAGCTGCCACCTTTGGCCAA 120
DB 1518 TGGAGGCGGAAATGAATGAATGGGTCTTCAGGAGACATGAGGGGTGCCAACCTTTGGCCAA 1577
QY 121 GCCCGCTCATGATCAACCGCTCTAAGAAAGACAGCGCTTGTCTCCCTGACGCGCGACC 180
DB 1578 GCGCTCTTGTGATTAAAGCACACTAAGAAAGATAGCCCTGCTTGTCTTGCACAGCTGACC 1637
QY 181 AGATGCTCAGTGCCTTTGTTGGATGCTGAGCGCCCATCTCTATTCCGAGTATGATCCTA 240
DB 1638 AGATGCTCAGTGCCTTTGTTGGATGCTGAGCGCCCATCTCTATTCTGAATATGATCCTT 1697
QY 241 CCAGACCCCTTCACTGAGAGCTTGTGATGATGGGCTTACTGACCAACCTGGCAGAGGAGC 300
DB 1698 CTAGACCCCTTCACTGAGAGCTTCAATGATGGGCTTATTGACCAACCTAGCAGATAGGAGC 1757
QY 301 TGGTTACATGATCAACTGGCGAGAGGGTGCAG 336
DB 1758 TGGTTTATATGATCAACTGGCGAGAGGGTGCAG 1793
RESULT 6
AK077236 2591 bp mRNA linear HTC 03-APR-2004
LOCUS
Mus musculus 11 days pregnant adult female ovary and uterus cDNA,
RIKEN full-length enriched library, clone:5031409B20
DEFINITION
product:estrogen receptor 1 (alpha), full insert sequence.
ACCESSION
AK077236.1 GI:26097248
VERSION
AK077236.1
KEYWORDS
HTC; CAP trapper.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1
Carninci,P. and Hayashizaki,Y.
AUTHORS
High-efficiency full-length cDNA cloning
TITLE
Meth. Enzymol. 303, 19-44 (1999)
JOURNAL
99279253
MEDLINE
10349636
PUBMED
2
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Sasaki,N., Carninci,P.,
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL
Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE
20499374
PUBMED
11042159
REFERENCE
AUTHORS
3
Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
Konno,H., Akiyama,J., Nishi,K., Kitsuunai,T., Tashiro,H., Itoh,M.,
Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsunura,S., Kawai,J.,
Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
JOURNAL
Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE
20530913

```


Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shingawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Takahashi, F., Takaku-Akaiura, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.

Direct Submission

Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.jp, URL: <http://genome.gsc.riken.jp/>, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Please visit our web site for further details.

URL: <http://genome.gsc.riken.jp/>

URL: <http://fantom.gsc.riken.jp/>

FEATURES	source	Location/Qualifiers
		1. .2735
		/organism="Mus musculus"
		/mol_type="mRNA"
		/strain="C57BL/6J"
		/db_xref="FANTOM_DB:E230025M07"
		/db_xref="taxon:10090"
		/clone="E230025M07"
		/sex="female"
		/tissue_type="oviduct"
		/clone_lib="RIKEN full-length enriched mouse cDNA library"
		/dev_stage="2 days pregnant adult"
		213_1713
misc_feature		/note="estrogen receptor 1 (alpha) (MGD MG1:1352467, GB NM_007956, evidence: BLASTN, 99%, match=1960) putative"

ORIGIN	Query Match	76.7%;	Score 257.6;	DB 3;	Length 2735;
	Best Local Similarity	85.4%;	Pred. No. 9.8e-62;		
	Matches 287;	Conservative 0;	Mismatches 49;	Indels 0;	Gaps 0;
Qy	1	GGATACCAAAAGACCGGAAGAGGAGGGGAGAATGTTGAAACAACAAGCGCCAGAGAGATGATG	60		
Db	986	GCATACGGAAAGACCGGAGGAGGGGAGAATGTTCAAGCACANGGTCAGAGAGATGACT	1045		
Qy	61	GGGAGGGCAGGGGTGAAGTGGGGTCTGCTGGAGACATGAGAGCTGCCAAACCTTTGGCCAA	120		
Db	1046	TGGAAAGCCGAAATGAAATGGTGTCTTCAGGAGACATGAGGGCTGCCAAACCTTTGGCCAA	1105		
Qy	121	GCCCGCTCATGATCAAAAGCTCTTAAGAAGACAGCGCTGGCCCTTGTCCTCGAGCGCCGACC	180		
Db	1106	GCCCTCTTGTGATTAAAGCACACTAAGAAGAAATAGCCCTGSCCTTGTCCTTGACAGCTGACC	1165		
Qy	181	AGATGGTCAGTGCCTTGTGGATGCTGAGCCCCCATACTCTATTCCGAGTATGATCCTTA	240		
Db	1166	AGATGGTCAGTGCCTTGTGGATGCTGAGCCCCCATGATCTATTCTGAAATGATCCTT	1221		
Qy	241	CCAGACCCCTTCAGTGAAGCTTCGATGATGGGCTTACTGACCAACCTGGCGAGCAGGGAGC	300		
Db	1.226	CTAGACCCCTTCAGTGAAGCCTCAATGATGGGCTTATTGACCAACCTAGCAGATAGGGAGC	1281		
Qy	301	TGGTTCATGATCAACTGGGGCGAAGAGGTCGCCAG	336		
Db	1286	TGGTTCATGATCAACTGGGGCGAAGAGAGTGCCAG	1321		

RESULT 8
AK041525
LOCUS
AK041525
4248 bp
mRNA
linear
HTC 03-APR-2004
Riken full-length enriched cDNA,
Mus musculus 3 days neonate thymus

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED

REFERENCE AUTHORS	TITLE	JOURNAL MEDLINE PUBMED
1. Kohn M, et al. (1996)
2. Smith J, et al. (1997)
3. Brown L, et al. (1998)
4. Davis R, et al. (1999)
5. Wilson T, et al. (2000)
6. Moore S, et al. (2001)
7. Taylor V, et al. (2002)
8. White H, et al. (2003)
9. Black G, et al. (2004)
10. Green I, et al. (2005)
11. Hall K, et al. (2006)
12. Young M, et al. (2007)
13. King N, et al. (2008)
14. Wright O, et al. (2009)
15. Scott P, et al. (2010)
16. Green Q, et al. (2011)
17. Adams R, et al. (2012)
18. Baker S, et al. (2013)
19. Nelson T, et al. (2014)
20. Carter U, et al. (2015)
21. Evans V, et al. (2016)
22. Foster W, et al. (2017)
23. Gibson X, et al. (2018)
24. Harlow Y, et al. (2019)
25. Jenkins Z, et al. (2020)
26. Kim A, et al. (2021)
27. Lee B, et al. (2022)
28. Miller C, et al. (2023)
29. Patel D, et al. (2024)
30. Roberts E, et al. (2025)

REFERENCE AUTHORS

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

TITLE
JOURNAL
REFERENCE
AUTHORS

TITLE
JOURNAL
REFERENCE
AUTHORS

TITLE
JOURNAL

COMMENT

library, clone:A630019F10 product:estrogen receptor 1 (alpha), full insert sequence.

AK041525
AK041525.1 GI:26334544
HTC; CAP trapper.
Mus musculus (house mouse)
Mus musculus

1. Carinci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
99279253
10349636

2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. *Genome Res.* 10 (10), 1617-1630 (2000)

20499374
11042159

3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuim, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwaga, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Wataniki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y. RIKEN integrated sequence analysis (RISA) system—384-format sequencing pipeline with 384 multicapillary sequencer Genome Res. 10 (11), 1757-1771 (2000)

20530913
11076861
4
The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
Genome of a full-length mouse cDNA collection

Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)
5 The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6 (bases 1 to 4248)

6 (basse 1 to 4248)
 Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
 Adachi, J., Aizawa, K., Akimura, T., Hara, A., Hashizume, W.,
 Fukuda, S., Furuno, M., Hanaigaki, T., Hara, A., Hashizume, W.,
 Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
 Hori, F., Imctani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
 Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
 Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
 Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ono, M., Ohsato, N.,
 Okazaki, Y., Saito, K., Satoh, H., Sakai, C., Sakai, K., Sakazume, N.,
 Sano, H., Sasaki, D., Shibata, K., Shiragawa, A., Shiraki, T.,
 Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
 Takeda, Y., Tanaka, T., Tomaru, A., Toyota, T., Yasunishi, A.,
 Muramatsu, M. and Hayashizaki, Y.

Direct Submission
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama
Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.jp,
URL: <http://genome.gsc.riken.jp/>, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Tissues were provided by Dr. John Todd (Dept. of Medical Genetics Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome Trust/WRC Building Addenbrookes Hospital Cambridge) whose assistance we gratefully acknowledge.
Please visit our web site for further details.
URL: <http://genome.gsc.riken.jp/>
URL: <http://fantom.gsc.riken.jp/>

FEATURES

```

source
1..4248
    /organism="Mus musculus"
    /mol_type="mRNA"
    /strain="C57BL/6J"
    /db_xref="FANTOM DB:A630019F10"
    /db_xref="taxon:10090"
    /clone="A630019F10"
    /tissue_type="thymus"
    /clone_lib="RIKEN full-length enriched mouse cDNA library"
    /dev_stage="3 days neonate"
    331..2130
    /note="unnamed protein product; estrogen receptor 1
(alpha) (MGI:1352467, GB|NM_007956, evidence: BLASTN,
99% match=1960)
putative"
    /codon_start=1
    /protein_id="BAC30973.1"
    /db_xref="GI:26334545"
    /translation="MTMTLHTKASGMALLHOIQNELEPLNRPOLKMPMERALGEVYV
DNSKPTVNYPEGAYENFAAARAAASAVYSGSGTAYGFGSEAAAFSANSAGFP
QLNSVSPFLMLHPGQVFLPHQYQVFIYLENEPSAVYRDTGPPAFYRNSDN
RQNGREPLKSNMIMESAKETQYCAVNDYAGHYGVMSCEGCAFFKRSIQ
GHNDYPTAQCTIDNRRKSCQACRLKRYEVMKGGIRKDRRGRLMKHRQD
DLEGRNEMASGDRAANLWPSPLVHKTKKNSPALSTADQVMSALDADPPMYSE
YDPSRPFSEASMGMLTNLADRELVMINWAKRVPFGDLNLHQVHLLECAULEIM
IGLVRSMEHPKQLFLAELLNDRNCKQVEGVEIFDMLLATSRFRMNLQGEFFV
CLKSIILNLSGVYFLSTLSKLEKHVRHLDKLTDLHLMAKAGLTQQQHRRL
AQLLIILSHIMNSKMEHLYNMKNVPLDYLLEMLDAHLRHAPASEMGVPEE
PSQTQATTSSTSAHSLQTYVYIPPEAEFPNTI"

```

ORIGIN

```

Query Match      76.7%; Score 257.6; DB 3; Length 4248;
Best Local Similarity 85.4%; Pred. No. 1.le-61;
Matches 287; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

Qy 1 CGATACGAAAGACCGAGAGAGGAGGAGAAATGTTGAAACACAGCGCAGAGAGATGATG 60
Db 1103 GCATACGGAAGACCGCGAGAGGAGGAGAAATGTTGAGCACACAGGCTCAGAGAGTACT 1162

Qy 61 GGGAGGGCAGGGGTGAAGTGGGGTCTGCTGGAGACATGAGAGCTGCCAACCTTTGGCCAA 120
Db 1163 TGGAGAGGCGGAAATGAAATGGGTGCTTCAGGAGACATGAGGCGTCCCAACCTTTGGCCAA 1222

Qy 121 GCGCGCTCATGATCAAAAGCTCTAAGAGAGACAGCTGCGCTTGTCCCTGACGCGCGACC 180
Db 1223 GCCCTCTTGATTAAGCACACTAAGAGAGATAGCCCTTGTCCCTTGTGACAGCTGACC 1282

Qy 181 AGATGGTCAGTCCCTTGTGGATGCTGAGCGCCGCCCATCTATTCCGAGTATGATCTCA 240
Db 1283 AGATGGTCAGTCCCTTGTGGATGCTGAGCGCCGCCCATCTATTCTGAATATGATCTTT 1342

Qy 241 CCAGACCTTCAGTGAAGCTTCGATGATGGGCTTACTGACCAACTGCGACACAGGGAGC 300
Db 1343 CTAGACCTTCAGTGAAGCTTCATGATGGGCTTATTGACCAACTGACAGATGGGAGC 1402

Qy 301 TGGTTCATCATCACTGCGGAGAGAGGTTGCCAG 336
Db 1403 TGGTTCATCATCACTGCGGAGAGAGTGGCCAG 1438

```

RESULT 9

```

AK087638
LOCUS      AK087638      4315 bp      mRNA      linear      HTC 03-APR-2004
DEFINITION Mus musculus 2 days pregnant adult female oviduct cDNA, RIKEN
full-length enriched library, clone:E230030D02 product:estrogen

```

ACCESSION

```

AK087638
VERSION    AK087638.1  GI:26104409
KEYWORDS   HTC; CAP trapper.
SOURCE     Mus musculus (house mouse)
ORGANISM   Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```

REFERENCE

```

AUTHORS    Carninci, P. and Hayashizaki, Y.
TITLE       High-efficiency full-length cDNA cloning
JOURNAL     Meth. Enzymol. 303, 19-44 (1999)
MEDLINE    99279253
PUBMED     10349636

```

REFERENCE

```

AUTHORS    Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE       Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL     Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE    20499374
PUBMED     11042159

```

REFERENCE

```

AUTHORS    Shibata, K., Itoh, M., Aizawa, K., Nagaoaka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M.,
Suni, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yanamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
TITLE       RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multiplexed sequencer
JOURNAL     Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE    20530913
PUBMED     11076861

```

REFERENCE

```

AUTHORS    The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
TITLE       Functional annotation of a full-length mouse cDNA collection
JOURNAL     Nature 409, 685-690 (2001)
MEDLINE    11076861

```

REFERENCE

```

AUTHORS    The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
TITLE       Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
JOURNAL     Nature 420, 563-573 (2002)
MEDLINE    11076861

```

REFERENCE

```

AUTHORS    Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
Muramatsu, M. and Hayashizaki, Y.
TITLE       Direct Submission
JOURNAL     Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.jp,
URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)

```

REFERENCE

```

AUTHORS    Muramatsu, M. and Hayashizaki, Y.
TITLE       CDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
JOURNAL     Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
COMMENT     Please visit our web site for further details.

```

```
FEATURES
  source
    URL:http://genome.gsc.riken.jp/
    URL:http://fantom.gsc.riken.jp/
  Location/Qualifiers
    1..4315
      /organism="Mus musculus"
      /mol_type="mRNA"
      /strain="C57BL/6J"
      /db_xref="FANTOM DB:E230030D02"
      /db_xref="taxon:10090"
      /clone="E230030D02"
      /sex="female"
      /tissue type="oviduct"
      /clone_lib="RIKEN full-length enriched mouse cDNA library"
      /dev_stage="2 days pregnant adult"
    1..4315
      /note="estrogen receptor 1 (alpha) (MGI:1352467,
      GB|NM_007956, evidence: BLASTN, 99%, match=1960)"

misc_feature
  1..4315
    /note="estrogen receptor 1 (alpha) (MGI:1352467,
    GB|NM_007956, evidence: BLASTN, 99%, match=1960)"

ORIGIN
  Query Match 76.7%; Score 257.6; DB 3; Length 4315;
  Best Local Similarity 85.4%; Pred.No. 1.1e-61;
  Matches 287; Conservative 0; Mismatches 49; Indels 0; Gaps 0;
  QY 1 GGATACGAAAGACCGAAGAGGAGGAGATGTTGAACACACAGCCGACGAGAGATGATG 60
  Db 935 GCATACGAAAGACCGCCGAGGAGGAGATGTTGAAGCACACAGCGTCAGAGAGATGACT 994
  QY 61 GCGAGGCGAGGGGTGAAGTGGGTCTGCTCGAGACATGAGAGCTGCCACCTTTGGCCAA 120
  Db 995 TGAAGGCGCGAAATGAATGAATGGGTCTTCAGGAGACATGAGGGCTGCCAACCTTTGGCCAA 1054
  QY 121 GCCCGCTCATGATCAACGCTCTAAGAAGAACGCTGGGCTTGTCCTGACGCGCGACC 180
  Db 1055 GCCTCTTTGATATAGCACACTAAGAGATAGCCCTGCTTGCTTGACAGCTGACC 1114
  QY 181 AGATGCTCAGTGCTTGTGGATGCTGAGCCCCCATCTCTATTTCGAGTATGATCCCTA 240
  Db 1115 AGATGCTCAGTGCTTGTGGATGCTGAGCCCCCATCTCTATTTCGATATGATCCCT 1174
  QY 241 CGAGACCTTCAGTGAAGCTTCGATGATGGCTTACTGACCACTCGGACAGAGGAGGC 300
  Db 1175 CTAGACCTTCAGTGAAGCTTCGATGATGGCTTATTGACCACTCGGACAGAGGAGGC 1234
  QY 301 TGGTTCACATGATCAACTGGCGCAAGAGGGTGCCAG 336
  Db 1235 TGGTTCATGATCAACTGGCGCAAGAGAGTGGCAG 1270

RESULT 10
AK036627
LOCUS
DEFINITION
  AK036627 4321 bp mRNA linear HTC 03-APR-2004
  Mus musculus adult male bone cDNA, RIKEN full-length enriched
  insert sequence.
ACCESSION
  AK036627.1 GI:26331559
VERSION
  AK036627.1
KEYWORDS
  HTC; CAP trapper.
SOURCE
  Mus musculus (house mouse)
  Mus musculus
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
  1
  Carninci, P. and Hayashizaki, Y.
  High-efficiency full-length cDNA cloning
  Meth. Enzymol. 303, 19-44 (1999)
JOURNAL
  95279253
MEDLINE
  10349636
PUBMED
  10349636
AUTHORS
  Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
  Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
  Normalization and subtraction of cap-trapper-selected cDNAs to
  prepare full-length cDNA libraries for rapid discovery of new genes
  Genome Res. 10 (10), 1617-1630 (2000)
JOURNAL
```

```
20499374
11042159
3
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
11076861
4
The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)
5
The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6 (bases 1 to 4321)
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
Hori, F., Inotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akai, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
Muramatsu, M. and Hayashizaki, Y.
Direct Submission
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.jp,
URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222,
Fax:81-45-503-9216)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL:http://genome.gsc.riken.jp/
URL:http://fantom.gsc.riken.jp/.
Location/Qualifiers
  1..4321
    /organism="Mus musculus"
    /mol_type="mRNA"
    /strain="C57BL/6J"
    /db_xref="FANTOM DB:9830143G19"
    /db_xref="taxon:10090"
    /clone="9830143G19"
    /sex="male"
    /tissue type="bone"
    /clone_lib="RIKEN full-length enriched mouse cDNA library"
    /dev_stage="adult"
  183..1982
    /note="unnamed protein product; estrogen receptor 1
    (alpha) (MGI:1352467, GB|NM_007956, evidence: BLASTN,
    99%, match=1960)
    putative"
    /codon_start=1

FEATURES
  source
    1..4321
      /organism="Mus musculus"
      /mol_type="mRNA"
      /strain="C57BL/6J"
      /db_xref="FANTOM DB:9830143G19"
      /db_xref="taxon:10090"
      /clone="9830143G19"
      /sex="male"
      /tissue type="bone"
      /clone_lib="RIKEN full-length enriched mouse cDNA library"
      /dev_stage="adult"
    183..1982
      /note="unnamed protein product; estrogen receptor 1
      (alpha) (MGI:1352467, GB|NM_007956, evidence: BLASTN,
      99%, match=1960)
      putative"
      /codon_start=1

CDS
```

```
/protein_id="BAC29510.1"
/db_xref="gi:26331560"
/translation="MTMTLTKRAGSMALLHQIGNELEPLNRPOLKMPMERALGEVYV
DNSKPTVFNYPGGAAYEFNAASAAAAASAPVYGSGIAYGSGSAAAFSANSGLAFP
OLNSVSPPLMLLHPPQLSPFLHPHQVPPYILENPSAYAVRTGCPFAFRKRSNQ
RQNGRRLSSNEKGNIMESAKETRYCAVNDYASGYHGWSCGCAFFKRSIQ
GHNDYMCATNCTIDKNRKRSCQARUKCYEVMGKGIKRRRGGRLKRRQD
DLEGNMGASGDMRAANLWPSPLVTKHTKNSPALSLTADQWVSGALLDAEPPMYSE
YDPSRPFSEASMMGLLTNLADRELVMINWAKRVPGFGLNLDHDOVHLLCAWLILM
IGLVWRSMEHPCKLLFAPNLLDRNCGKCEGVRIFFMLLATSSRRMNLQGBEFV
CLKSLILLNSGVYTELSSFLKSLERKHHRVLDKLTDLHLMAKAGLTLQOQRRL
QQLLILSHIRMSKNGEHLNMMCKKNVPLYDLLEMLDAHRLHAPASRMGVPPPE
ESQQLATSTSAHSLQTYIYPPEAGFPNTI"
polyA_signal
4306..4311
/note="putative"
polyA_site
4321
/note="putative"
```

ORIGIN

```
Query Match 76.7%; Score 257.6; DB 3; Length 4321;
Best Local Similarity 85.4%; Pred. No. 1.1e-61;
Matches 287; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 1 GGATACGAAAGACCGAAGAGAGAGGAGATGTTGAACACAAAGCGCCAGAGATGATG 60
Db |||||
QY 955 GCATACGAAAGACCGCGAGAGAGGAGATGTTGAAGCACAAGCGTCAGAGATGACT 1014
Db |||||
QY 61 GGGAGGCGAGGGGTGAAGTGGGGTCTGCTGGAGACATGAGCTGCCAACCTTTGGCCAA 120
Db |||||
QY 1015 TGGAGGCGGAATGAAATGGGTCTTCAGGAGATGAGGCTGCCAACCTTTGGCCAA 1074
Db |||||
QY 121 GCGGCTCATGATCAACCGCTCTAAGAGAACAGCGCTGGCCCTTGCTGACGCGGACC 180
Db |||||
QY 1075 GCGCTCTTGTGATTAAGCACACTAAGAGAAATAGCCCTTGCTTGACAGCTGACC 1134
Db |||||
QY 181 AGATGGTCAGTCCCTTGTGATGCTGAGCCCCCATCTCTATTCCGAGTATGATCTTA 240
Db |||||
QY 1135 AGATGGTCAGTCCCTTGTGATGCTGAACCGCCCATGATCTATTCTGAATATGATCTT 1194
Db |||||
QY 241 CCAGACCTTCAGTGAAGCTTCGATGATGGGCTTACTGACCAACCTGCAGACAGGAGC 300
Db |||||
QY 1195 CTAGACCTTCAGTGAAGCTCAATGATGGGCTTATTGACCAACCTAGCAGATGGAGC 1254
Db |||||
QY 301 TGGTTACATGATCAACTGGGCGAAGAGGGTGCCAG 336
Db |||||
QY 1255 TGGTTACATGATCAACTGGGCGAAGAGAGTGCCAG 1290
Db |||||
```

RESULT 11

```
AA023625/c
LOCUS
DEFINITION
  mh80b06.r1 Soares mouse placenta 4NbMP13.5 14.5 Mus musculus cDNA
  clone IMAGE:457235 5' similar to gb:M12674 ESTROGEN RECEPTOR
  (HUMAN); gb:M38651 Mouse estrogen receptor mRNA, complete cds
  (MOUSE);, mRNA sequence.
ACCESSION
  AA023625
VERSION
  AA023625.1 GI:1487542
KEYWORDS
  EST.
SOURCE
  Mus musculus (house mouse)
```

ORGANISM

```
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 593)
Matra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
```

TITLE

The WashU-HHMI Mouse EST Project

JOURNAL

Unpublished (1996)

COMMENT

Contact: Marra M/Mouse EST Project

WashU-HHMI Mouse EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LML; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:274123

Seq primer: -28M13 rev2 from Amersham

High quality sequence stop: 492.

Location/Qualifiers

1..593

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="IMAGE:457235"

/sex="unknown"

/tissue type="placenta"

/dev stage="adult"

/lab_host="DH10B"

/clone_lib="Soares mouse placenta 4NbMP13.5 14.5"

/note="Organ: placenta; Vector: p7T3D-Pac (Pharmacia)

with a modified polylinker; Site 1: Not 1; Site 2: Eco RI;

1st strand cDNA was primed with a Not I - oligo(dT) primer

15,

TGTTACCAATCTGAAGTCGGAGCGCGGGAATTTTTTTTTTTTTTTTTTTT

T 3'; double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia), digested with Not I and cloned into the Not I

and Eco RI sites of the modified p7T3 vector. Library

went through one round of normalization, and was

constructed by Bento Soares and M.Fatima Bonaldo."

ORIGIN

```
Query Match 73.1%; Score 245.6; DB 1; Length 593;
Best Local Similarity 85.1%; Pred. No. 1.8e-58;
Matches 286; Conservative 0; Mismatches 49; Indels 1; Gaps 1;

QY 1 GGATACGAAAGACCGAAGAGAGGAGAAATGTTGAACACAAAGCGCCAGAGATGATG 60
Db |||||
QY 499 GCATACGAAAGACCGCGGAGAGGAGAAATGTTGAAGCAAGCGTCAGAGATGACT 440
Db |||||
QY 61 GGGAGGCGAGGGGTGAAGTGGGGTCTGCTGGAGACATGAGAGCTGCCAACCTTTGGCCAA 120
Db |||||
QY 439 TGGAGGCGGGAATGAATGGTGTCTTCAGAGACATGAGGCTGCCAACCTTT-GCCAA 381
Db |||||
QY 121 GCGGCTCATGATCAACCGCTCTAAGAGAACAGCGCTGGCCCTTGCTGACGCGGACC 180
Db |||||
QY 380 GCGCTCTTGTGATTAAGCACACTAAGAGAAATAGCCCTTGCTCTTGACAGCTGACC 321
Db |||||
QY 181 AGATGGTCAGTCCCTTGTGATGCTGAGCCCCCATCTCTATTCCGAGTATGATCTTA 240
Db |||||
QY 320 AGATGGTCAGTCCCTTGTGATGCTGAGCCCCCATCTCTATTCTGAATATGATCTT 261
Db |||||
QY 241 CCAGACCTTCAGTGAAGCTTCGATGATGGGCTTACTGACCAACCTGCAGACAGGAGC 300
Db |||||
QY 260 CTAGACCTTCAGTGAAGCTTCATGATGGGCTTATTGACCAACCTAGACAGTAGGGAGC 201
Db |||||
QY 301 TGGTTACATGATCAACTGGGCGAAGAGGGTGCCAG 336
Db |||||
QY 200 TGGTTACATGATCAACTGGGCGAAGAGAGTGCCAG 165
Db |||||
```

RESULT 12

```
AA092184
LOCUS
DEFINITION
  114609.seq.F Human fetal heart, Lambda ZAP Express Homo sapiens
  cDNA 5', mRNA sequence.
ACCESSION
  AA092184
VERSION
  AA092184.1 GI:1637173
KEYWORDS
  EST.
  Homo sapiens (human)
SOURCE
  ORGANISM
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```



```

QY 29 AATGTTGAACACAGCG-CCAGAGAGATGATGGGAGGCGCAGGGTGAGTGGGGCTG 87
Db 2 AATGTTTAAGCCCAAGCGTCAAAAAGATAAATCTGGAAGGCCAAATGAATGGGTGCT 61
QY 88 CTGG-AGACATGAGAGCTGCCACCTTGGCCNAAGCCGCTCATGATCAACGCTCTAAG 146
Db 62 CAGGAAGACATGAGGGTTGCCAACCTTTGGCCAAAGCCCTCTTGTGATTAAAGCACACTAAG 121
QY 147 AAGAACAGAGCTGCGCTTGTCCCTGACGGCGGACAGATGGTCAAGTCCCTTGTGGATGCT 206
Db 122 AAGAATAGCCCTGCTTGTCTTGTGAGCTGACCATGCTTGTGATGCT 181
QY 207 GAGCCCCCATCTACTTATTCGAGTAGATGATCTTACAGACCTTCAAGCTTCGATG 266
Db 182 GAACGCGCCCATGATCTTATCTGAATATGATCTTCTAGACCTTCAAGTGAAGCCTCAAG 241
QY 267 ATGGGCTTACTGACCAACCTGCGACAGAGGAGCTGTTCAATGATCAACTGGGCGAAG 326
Db 242 ATGGGCTTATTGACCAACCTAGCAGATAGGAGCTGGTTTCATATGATCAACTGGGCAAG 301
QY 327 AGGTTGCCAG 336
Db 302 AGAGTGCCAG 311

RESULT 14
AU041270/c
LOCUS
DEFINITION
AU041270 Mouse four-cell-embryo cDNA Mus musculus cDNA clone
J0828F12.3', mRNA sequence.
ACCESSION
AU041270
VERSION
AU041270.1 GI:3955505
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus

REFERENCE
1 (bases 1 to 576)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Ko.M.S.H., Kitchen,J.R., Wang,X., Threat,T.A., Sun,T.,
DePalma,G.E., Liang,Y., Kargul,G.J., Sharara,R., Lim,M.K. and
Doi,H.
Systematic analyses of genes expressed in 4-cell mouse embryo (The
ERATO/Doi Project at Wayne State University)
Unpublished (1998)
Contact: Hirofumi Doi
Doi Biosymmetry Project, ERATO
Japan Science and Technology Corporation (JST)
WBG-Marive East 12F, 2-6 Nakase, Mihama-ku, Chiba 261-71, Japan
Email: hdbioa.jst.go.jp.
Location/Qualifiers
1..576
/organism="Mus musculus"
/mol_type="mRNA"
/strain="CS7BL/6J"
/db_xref="taxon:10090"
/clone="J0828F12"
/dev_stage="four-cell-embryo"
/clone_lib="Mouse four-cell-embryo cDNA"

ORIGIN
Query Match 50.8%; Score 170.8; DB 1; Length 576;
Best Local Similarity 81.6%; Pred. No. 2.6e-37;
Matches 209; Conservative 0; Mismatches 44; Indels 3; Gaps 1;

QY 81 GGGTCTGCTGAGACATGAGAGCTGCCAACCTTTGGCCAAAGCCCTCATGATCAAGCGC 140
Db 571 GGGNCCTTCGGAGACATGAGGCTGCCAACCTTTGGCCAAAGCCCTCTTGTGATTA--GC 515
QY 141 TCTAAGAAGAACAGCGCTGGCTTGTCCCTGAGCGGCGGACCATGCTGCTGCTTGTG 200
Db 514 ACACATAGAGATACCCCTTGTTCCTTAAACAGCTGACCATGCTGCTGCTTGTG 455
QY 201 GATGCTGAGCCCCCATACTACTTATTCGAGTAGATGATCTTACAGACCCCTTCAGTGAAGCT 260

```

```

Db 454 AATGCTGAACCGCCCATGATCTATTCTGATATGATCTCTAGACCCCTTCAGTGAAGCC 395
QY 261 TCATGATGGCTTACTGACCAACCTGGCAGACAGGAGCTGGTTCAATGATCAACTGG 320
Db 394 TCAATGATGGCTTATTGACCAACCTAGCAGATAGGAGCTGGTTTCATATGATCAACTGG 335
QY 321 GCGAAGAGGGTGCCAG 336
Db 334 GCARAAGAGAGTGCCAG 319

BB097271 316 bp mRNA linear EST 26-JUN-2000
BB097271 RIKEN full-length enriched, 12 days embryo, embryonic body
between diaphragm region and neck Mus musculus cDNA clone
9430058N13.3', mRNA sequence.
ACCESSION
BB097271
VERSION
BB097271.1 GI:8742305
KEYWORDS
EST.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus

REFERENCE
1 (bases 1 to 316)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T.,
Carninci,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,
Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,
Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C.,
Kusakabe,M., Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H.,
Okazaki,Y., Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K.,
Shibata,K., Shibata,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T.,
Sogabe,Y., Sugahara,Y., Suzuki,H., Suzuki,H., Tagawa,A.,
Takahashi,F., Tomimaga,N., Toya,T., Tsunoda,Y., Watahiki,A.,
Watanabe,S., Yamamura,T., Yamanaoka,I., Yano,R., Yasunishi,A.,
Yokota,T., Yoshida,K., Yoshiki,A., Yoshino,M., Muramatsu,M. and
Hayashizaki,Y.
RIKEN Mouse ESTs (Konno,H., et al.)
Unpublished (2000)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suenho-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoaka,S.,
Sasaki,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Thermolabile and thermoactivation of thermolabile enzymes by
trehalose and its application for the synthesis of full length
cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
Tonari,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M.,
Okazaki,Y. and Hayashizaki,Y.
Automated filtration-based high-throughput plasmid preparation
system. Genome Res. 9 (5), 463-470 (1999)
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for
further details.
Location/Qualifiers
1..316
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="9430058N13"
/tissue_type="embryonic body between diaphragm region and
neck"
/dev_stage="12 days embryo"

FEATURES
source
1..316

```

```
/lab_host="DH10B"
/clone_lib="RIKEN full-length enriched, 12 days embryo,
embryonic body between diaphragm region and neck"
/note="Site 1: Sali; Site 2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5',
GAGAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot = 20.0 and subtraction to Rot = 370.0. Second
strand cDNA was prepared with the primer adapter of
sequence [5', GAGAGAGAGATCTCGAGTTAATAATTAATCCCCCCCCCCC
3']. cDNA was cleaved with XhoI and BamHI."
```

ORIGIN

```
Query Match      35.2%; Score 118.4; DB 2; Length 316;
Best Local Similarity 83.8%; Pred. No. 1.5e-22;
Matches 134; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

Qy      177  GACCAGATGTCAGTGCTTTGTTGGATGCTGAGCCCCCATCTCTATTCCGAGTATGAT 236
Db      2    ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy      237  CCTACCAGACCTTCAGTGAAGCTTCGATGATGGGCTTACTGACCAACCTGGCAGACAGG 296
Db      62  ||||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy      297  GAGCTGGTTTCATGATCACTGGGCGAAGAGGGTGCCAG 336
Db      122  GAGCTGGTTTATATGATCACTGGGCGAAGAGAGTGCCAG 161
```

Search completed: November 3, 2004, 09:43:36
Job time : 5803.75 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model
Run on: November 3, 2004, 04:55:34 ; Search time 2096.38 Seconds
(without alignments)
7579.407 Million cell updates/sec

Title: US-09-933-267A-1_COPY_306168_306503
Perfect score: 336
Sequence: 1 ggatcgaaaagaccgaaga.....ctggggcgaagaggggtgcag 336

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4526729 seqs, 2364484745 residues
Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Score	Query Match Length DB ID	Description
1	336	100.0 1223 6	AX066401 Sequence
2	336	100.0 1237 9	AF123496 Homo sapi
3	336	100.0 1374 6	AX066402 Sequence
4	336	100.0 1770 6	AX824420 Sequence
5	336	100.0 1788 6	AX411751 Sequence
6	336	100.0 1803 6	AX824417 Sequence
7	336	100.0 2092 6	I08538 Sequence 1
8	336	100.0 2092 6	AX474709 Sequence
9	336	100.0 2092 9	HUMERMCF
10	336	100.0 2106 9	HSU47678
11	336	100.0 2220 6	B0016996 Fused est
12	336	100.0 2322 6	A42099 Sequence 1
13	336	100.0 2532 6	BD187708 Specific
14	336	100.0 5439 9	HSM807087
15	336	100.0 6450 6	AR153585 Sequence
16	336	100.0 6450 6	AR225684 Sequence
17	336	100.0 6450 6	AR397400 Sequence
18	336	100.0 6450 6	AX587952 Sequence
19	336	100.0 6450 9	HSERR

20	336	100.0	6610	12	AF061181	Mammalian
21	336	100.0	6833	6	AX128349	Sequence
22	336	100.0	6956	6	AX128348	Sequence
23	336	100.0	7038	6	AX128342	Sequence
24	336	100.0	11365	6	AX512493	Sequence
25	336	100.0	165237	6	AX232505	Sequence
26	336	100.0	299637	9	AX425004	Homo sapi
27	336	100.0	349980	6	AX232503	Sequence
28	336	100.0	349980	6	AX453703	Sequence
29	334.4	99.5	139687	9	AL590993	Human DNA
30	333	99.1	8566	6	AR397407	Sequence
31	332.8	99.0	4963	6	AR029418	Sequence
32	332.8	99.0	4963	6	I15368	Sequence 16
33	332.8	99.0	4963	6	I56760	Sequence 9
34	325	96.7	6639	6	AX128351	Sequence
35	325	96.7	6695	6	AX128347	Sequence
36	325	96.7	6695	6	AX128353	Sequence
37	325	96.7	6695	6	AX128354	Sequence
38	325	96.7	6801	6	AX128355	Sequence
39	325	96.7	6818	6	AX128346	Sequence
40	325	96.7	6900	6	AX128341	Sequence
41	320.4	95.4	392	9	AF120105	Homo sapi
42	288	85.7	1788	4	SSESTREC	Sequence
43	280	83.3	1966	4	OAESTREC	Sequence
44	280	83.3	2327	4	AY538775	Bos tauru
45	267.2	79.5	3746	4	AF124093	Equus cab

ALIGNMENTS

RESULT 1
AX066401
LOCUS AX066401 1223 bp DNA linear PAT 24-JAN-2001
DEFINITION Sequence 1 from Patent WO0100823.
ACCESSION AX066401
VERSION AX066401.1 GI:12544111
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Gannon, F., Denger, S. and Flouriot, G.
TITLE Novel isoforms of the human estrogen receptor-(g(a)
JOURNAL Patent: WO 010823-A 1 04-JAN-2001;
EUROPEAN MOLECULAR BIOLOGY LABORATORY (DE)
FEATURES
source
1. 1223
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN			
Query Match	100.0%	Score 336;	DB 6; Length 1223;
Best Local Similarity	100.0%;	Pred. No. 2.1e-85;	
Matches 336;	Conservative 0;	Mismatches 0;	Indels 0; Gaps 0;
Qy	1	GGATACGAAAGACCGAAGAGGGAGAGATGTTGAACACAAAGCCGACAGAGATGATG 60	
Db	196	GGATACGAAAGACCGAAGAGGGAGAGATGTTGAACACAAAGCCGACAGAGATGATG 255	
Qy	61	GGGAGGCGAGGGGTGAAGTGGGTCTGCTGGAGACATGAGAGCTGCCAAGCTTGGCCAA 120	
Db	256	GGGAGGCGAGGGGTGAAGTGGGTCTGCTGGAGACATGAGAGCTGCCAAGCTTGGCCAA 315	
Qy	121	GCCCGCTCATGATCAACCGCTCTAAGAAAGAACAGCCCTTGCCCTTGCACGCCGAC 180	
Db	316	GCCCGCTCATGATCAACCGCTCTAAGAAAGAACAGCCCTTGCCCTTGCACGCCGAC 375	
Qy	181	AGATGGTCATGCTGTTGGATGCTGAGCCCCCCTACTCTCTATTCGAGATGATGATCCTA 240	
Db	376	AGATGGTCATGCTGTTGGATGCTGAGCCCCCCTACTCTCTATTCGAGATGATGATCCTA 435	

```
QY 241 CCAGACCCCTTCAGTGAAGCTTCGATGATGGCTTACTGACCAACCTGGCAGACAGGGAGC 300
Db 436 CCAGACCCCTTCAGTGAAGCTTCGATGATGGCTTACTGACCAACCTGGCAGACAGGGAGC 495
QY 301 TGGTTACATGATCAACTGGCGAAGAGGGTGCAG 336
Db 496 TGGTTACATGATCAACTGGCGAAGAGGGTGCAG 531

RESULT 2
HSESRI3
LOCUS Homo sapiens estrogen receptor alpha (ESR1) gene, exon 4.
DEFINITION AF123496
ACCESSION AF123496
VERSION AF123496.1 GI:5821720
KEYWORDS
SEGMENT
SOURCE
ORGANISM Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1237)
AUTHORS Schubert,E.L., Lee,M.K., Newman,B. and King,M.C.
TITLE Single nucleotide polymorphisms (SNPs) in the estrogen receptor
JOURNAL gene and breast cancer susceptibility
MEDLINE J. Steroid Biochem. Mol. Biol. 71 (1-2), 21-27 (1999)
20084372
FUBMED 10619354
REFERENCE 2 (bases 1 to 1237)
AUTHORS Schubert,E.L., Lee,M.K. and King,M.-C.
TITLE Direct Submission
JOURNAL Submitted (26-JAN-1999) Medical Genetics, University of Washington,
1959 NE Pacific Street, Seattle, WA 98195-7720, USA
FEATURES
source
Location/Qualifiers
1..1237
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chrnosome="6"
/map="6q25"
350..685
/genes="ESR1"
/number=4
exon
ORIGIN
Query Match 100.0%; Score 336; DB 9; Length 1237;
Best Local Similarity 100.0%; Pred. No. 2.1e-85;
Matches 336; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GGATACGAAAGACCGAAGAGGAGGAGAGTGTGAAACACACAGCGCCAGAGAGATGATG 60
Db 350 GGATACGAAAGACCGAAGAGGAGGAGAGTGTGAAACACACAGCGCCAGAGAGATG 409
QY 61 GGGAGGGCGAGGGGTGAAGTGGGGTCTGCTGGAGACATGAGCTGCCAACCTTTGGCCAA 120
Db 410 GGGAGGGCGAGGGGTGAAGTGGGGTCTGCTGGAGACATGAGCTGCCAACCTTTGGCCAA 469
QY 121 GCCCGCTCATGATCAACCGCTCTAAGAGACAGCTGGCTTGTCCCTTGACGGCGGACC 180
Db 470 GCCCGCTCATGATCAACCGCTCTAAGAGACAGCTGGCTTGTCCCTTGACGGCGGACC 529
QY 181 AGATGCTCAGTGCCTTGTGGATGCTGAGCGCCCGCCATCTCTATTCGAGTATGATCCTA 240
Db 530 AGATGCTCAGTGCCTTGTGGATGCTGAGCGCCCGCCATCTCTATTCGAGTATGATCCTA 589
QY 241 CCAGACCCCTTCAGTGAAGCTTCGATGATGGCTTACTGACCAACCTGGCAGACAGGGAGC 300
Db 590 CCAGACCCCTTCAGTGAAGCTTCGATGATGGCTTACTGACCAACCTGGCAGACAGGGAGC 649
QY 301 TGGTTACATGATCAACTGGCGAAGAGGGTGCAG 336
Db 650 TGGTTACATGATCAACTGGCGAAGAGGGTGCAG 685
```

```
RESULT 3
AX066402
LOCUS Homo sapiens estrogen receptor alpha (ESR1) gene, exon 4.
DEFINITION AX066402
ACCESSION AX066402
VERSION AX066402.1 GI:12544112
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Gannon,F., Denger,S. and Flourot,G.
TITLE Novel isoforms of the human estrogen receptor-g(a)
JOURNAL Patent: WO 0100823-A 2 04-JAN-2001;
EUROPEAN MOLECULAR BIOLOGY LABORATORY (DE)
FEATURES
source
Location/Qualifiers
1..1374
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 100.0%; Score 336; DB 6; Length 1374;
Best Local Similarity 100.0%; Pred. No. 2.1e-85;
Matches 336; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GGATACGAAAGACCGAAGAGGAGGAGAGTGTGAAACACACAGCGCCAGAGATGATG 60
Db 347 GGATACGAAAGACCGAAGAGGAGGAGAGTGTGAAACACACAGCGCCAGAGATGATG 406
QY 61 GGGAGGGCGAGGGGTGAAGTGGGGTCTGCTGGAGACATGAGCTGCCAACCTTTGGCCAA 120
Db 407 GGGAGGGCGAGGGGTGAAGTGGGGTCTGCTGGAGACATGAGCTGCCAACCTTTGGCCAA 466
QY 121 GCCCGCTCATGATCAACCGCTCTAAGAGACAGCTGGCTTGTCCCTTGACGGCGGACC 180
Db 467 GCCCGCTCATGATCAACCGCTCTAAGAGACAGCTGGCTTGTCCCTTGACGGCGGACC 526
QY 181 AGATGCTCAGTGCCTTGTGGATGCTGAGCGCCCGCCATCTCTATTCGAGTATGATCCTA 240
Db 527 AGATGCTCAGTGCCTTGTGGATGCTGAGCGCCCGCCATCTCTATTCGAGTATGATCCTA 586
QY 241 CCAGACCCCTTCAGTGAAGCTTCGATGATGGCTTACTGACCAACCTGGCAGACAGGGAGC 300
Db 587 CCAGACCCCTTCAGTGAAGCTTCGATGATGGCTTACTGACCAACCTGGCAGACAGGGAGC 646
QY 301 TGGTTACATGATCAACTGGCGAAGAGGGTGCAG 336
Db 647 TGGTTACATGATCAACTGGCGAAGAGGGTGCAG 682

RESULT 4
AX824420
LOCUS An estrogen receptor alpha construct for use in a yeast 2-hybrid
DEFINITION AX824420
ACCESSION AX824420
VERSION AX824420.1 GI:39750420
KEYWORDS
SOURCE synthetic construct
ORGANISM synthetic construct
artificial sequences.
REFERENCE 1
AUTHORS Verdonk,G.N., Dijkema,R.N. and Schoonen,W.G.
TITLE An estrogen receptor alpha construct for use in a yeast 2-hybrid
JOURNAL assay
Patent: WO 03070975-A 6 28-AUG-2003;
Akzo Nobel N.V. (NL)
FEATURES
source
Location/Qualifiers
1..1770
/organism="synthetic construct"
```

```
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="Fusion DNA sequence page 17-19"

ORIGIN

Query Match      100.0%; Score 336; DB 6; Length 1770;
Best Local Similarity 100.0%; Pred. No. 2.1e-85;
Matches 336; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  GGATACGAAAGACCGAAGAGGAGGAGGAGATGTTGAAACACACAGCGCCAGAGAGATGATG 60
Db      743 GGATACGAAAGACCGAAGAGGAGGAGGAGATGTTGAAACACACAGCGCCAGAGAGATGATG 802

QY      61  GGGAGGCGCAGGGGTGAAGTGGGCTCTCTGGAGACATGAGAGCTGCGCAACTTTGGCCAA 120
Db      803 GGGAGGCGCAGGGGTGAAGTGGGCTCTCTGGAGACATGAGAGCTGCGCAACTTTGGCCAA 862

QY      121 GCCCGCTCATGATCAAAACGCTCTAAGAGAACAGCCTGGCGCTTGCCCTGACGCGCGACC 180
Db      863 GCCCGCTCATGATCAAAACGCTCTAAGAGAACAGCCTGGCGCTTGCCCTGACGCGCGACC 922

QY      181 AGATGTTCAGTGCCTTGTGGATGCTGAGCCCCCATCTCTATTCCGAGTATGATCCTA 240
Db      923 AGATGTTCAGTGCCTTGTGGATGCTGAGCCCCCATCTCTATTCCGAGTATGATCCTA 982

QY      241 CCAGACCCCTTCAGTGAAGCTTCGATGATGGCTTACTGACCAACCTGGCAGACAGGAGC 300
Db      983 CCAGACCCCTTCAGTGAAGCTTCGATGATGGCTTACTGACCAACCTGGCAGACAGGAGC 1042

QY      301 TGGTTCCATGATCAACTGGCGCAAGAGGTGCCAG 336
Db      1043 TGGTTCCATGATCAACTGGCGCAAGAGGTGCCAG 1078

RESULT 5
AX411751
LOCUS      AX411751      1788 bp      DNA      linear      PAT 14-JUN-2002
DEFINITION Sequence 1 from Patent WO0228175.
ACCESSION  AX411751
VERSION     AX411751.1 GI:21444272
KEYWORDS
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
            Chambon,P. and Metzger,D.
            Transgenic mouse for targeted recombination mediated by modified
            Cre-er
            Patent: WO 0228175-A 1 11-APR-2002;
            ASS POUR LE DEV DE LA RECH (FR)
FEATURES
            Location/Qualifiers
                1..1788
                /organism="Homo sapiens"
                /mol_type="unassigned DNA"
                /db_xref="taxon:9606"
                1..1788
                /note="unnamed protein product"
                /codon_start=1
                /protein_id="CAD35258.1"
                /db_xref="GI:21444273"
                /translation="MTMTLTKASGMALLHQIQGNEPLNRPOLKIPLERPLGEVYL
                DSKPAVNPPEGAAYEFNAAANAAVYGTGLPGSEAAAFGNSLGLGFPPLNS
                VSPFLMLLHPPLQSPFLPHGQVPPVYLENEPSGYTVREAGPAPFVPPNSDRQ
                GPERLASTNDKGSMAWESAKETRYCANCNDYASGHYVWCEGKAPFKRSIQGHND
                YMCPTNQCTIDNRRKSQCARLRCXCEYVGMWKGIRKDRRGKMLKHKRQRDDGEG
                REVSAGSDMRAANLWPSLMTKRSKNLSLSTADQVSLDAEPPLIYSEYDPT
                REFSASMLNLADRLVHMINKRVPQFVLTLDHDOVHLLCEAWLEILMIGLV
                WRSMEHPGLLPAPLLDNRNOKCVGWEVIFDMLATSSRFMMNLQGESEFVLKS
                IILLNSGVYTFSLSLLEEKDHHRVLDKTDILHLIMAKAGLTLQQOHRLAQLL
                LILSHRMSNKGMEHLYSMCKNVVFLDLELLELLELDHRAHLHAPTSRGGASVEETDQS
                HLATAGSTSSLSLQKYYITGEABGFPATV"

ORIGIN
```

```
Query Match      100.0%; Score 336; DB 6; Length 1788;
Best Local Similarity 100.0%; Pred. No. 2.1e-85;
Matches 336; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  GGATACGAAAGACCGAAGAGGAGGAGGAGATGTTGAAACACACAGCGCCAGAGAGATGATG 60
Db      761 GGATACGAAAGACCGAAGAGGAGGAGGAGATGTTGAAACACACAGCGCCAGAGAGATGATG 820

QY      61  GGGAGGCGCAGGGGTGAAGTGGGCTCTCTGGAGACATGAGAGCTGCGCAACTTTGGCCAA 120
Db      821 GGGAGGCGCAGGGGTGAAGTGGGCTCTCTGGAGACATGAGAGCTGCGCAACTTTGGCCAA 880

QY      121 GCCCGCTCATGATCAAAACGCTCTAAGAGAACAGCCTGGCGCTTGCCCTGACGCGCGACC 180
Db      881 GCCCGCTCATGATCAAAACGCTCTAAGAGAACAGCCTGGCGCTTGCCCTGACGCGCGACC 940

QY      181 AGATGTTCAGTGCCTTGTGGATGCTGAGCCCCCATCTCTATTCCGAGTATGATCCTA 240
Db      941 AGATGTTCAGTGCCTTGTGGATGCTGAGCCCCCATCTCTATTCCGAGTATGATCCTA 1000

QY      241 CCAGACCCCTTCAGTGAAGCTTCGATGATGGCTTACTGACCAACCTGGCAGACAGGAGC 300
Db      1001 CCAGACCCCTTCAGTGAAGCTTCGATGATGGCTTACTGACCAACCTGGCAGACAGGAGC 1060

QY      301 TGGTTCCATGATCAACTGGCGCAAGAGGTGCCAG 336
Db      1061 TGGTTCCATGATCAACTGGCGCAAGAGGTGCCAG 1096

RESULT 6
AX824417
LOCUS      AX824417      1803 bp      DNA      linear      PAT 11-DEC-2003
DEFINITION Sequence 3 from Patent WO03070975.
ACCESSION  AX824417
VERSION     AX824417.1 GI:39750419
KEYWORDS    synthetic construct
            synthetic construct
            artificial sequences.
SOURCE
ORGANISM    Verdonk,G.N., Dijkema,R.N. and Schoonen,W.G.
            An estrogen receptor alpha construct for use in a yeast 2-hybrid
            assay
            Patent: WO 03070975-A 3 28-AUG-2003;
            Akzo Nobel N.V. (NL)
FEATURES
            Location/Qualifiers
                1..1803
                /organism="synthetic construct"
                /mol_type="unassigned DNA"
                /db_xref="taxon:32630"
                /note="Fusion DNA p 14-16"

ORIGIN

Query Match      100.0%; Score 336; DB 6; Length 1803;
Best Local Similarity 100.0%; Pred. No. 2.1e-85;
Matches 336; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  GGATACGAAAGACCGAAGAGGAGGAGGAGATGTTGAAACACACAGCGCCAGAGAGATGATG 60
Db      776 GGATACGAAAGACCGAAGAGGAGGAGGAGATGTTGAAACACACAGCGCCAGAGAGATGATG 835

QY      61  GGGAGGCGCAGGGGTGAAGTGGGCTCTCTGGAGACATGAGAGCTGCGCAACTTTGGCCAA 120
Db      836 GGGAGGCGCAGGGGTGAAGTGGGCTCTCTGGAGACATGAGAGCTGCGCAACTTTGGCCAA 895

QY      121 GCCCGCTCATGATCAAAACGCTCTAAGAGAACAGCCTGGCGCTTGCCCTGACGCGCGACC 180
Db      896 GCCCGCTCATGATCAAAACGCTCTAAGAGAACAGCCTGGCGCTTGCCCTGACGCGCGACC 955

QY      181 AGATGTTCAGTGCCTTGTGGATGCTGAGCCCCCATCTCTATTCCGAGTATGATCCTA 240
Db      956 AGATGTTCAGTGCCTTGTGGATGCTGAGCCCCCATCTCTATTCCGAGTATGATCCTA 1015
```

QY 241 CCAGACCTTCAGTGAAGCTTCGATGATGGCTTACTGACCAACCTGGCAGACGGGAGC 300
Db 1016 CCAGACCTTCAGTGAAGCTTCGATGATGGCTTACTGACCAACCTGGCAGACGGGAGC 1075
QY 301 TGGTTTCATGATCAACTGGGCGAAGAGGGTGCCAG 336
Db 1076 TGGTTTCATGATCAACTGGGCGAAGAGGGTGCCAG 1111

RESULT 7
LOCUS I08538 108538 2092 bp DNA linear PAT 02-DEC-1994
DEFINITION Sequence 1 from Patent WO 8705049.
ACCESSION I08538
VERSION I08538.1 GI:588753
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
UNCLASSIFIED

REFERENCE 1 (bases 1 to 2092)
AUTHORS Shine,J.
TITLE EUCARYOTIC EXPRESSION OF STEROID RECEPTOR PROTEINS
JOURNAL Patent: WO 8705049-A 1 27-AUG-1987;
FEATURES Location/Qualifiers
source 1..2092
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN

Query Match 100.0%; Score 336; DB 6; Length 2092;
Best Local Similarity 100.0%; Pred. No. 2.1e-85;
Matches 336; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GGATACGAAAGACCGAAGAGAGGAGGAGAGATGTTGAACACAAAGCGCCAGAGATGATG 60
Db 1053 GGATACGAAAGACCGAAGAGAGGAGGAGAGATGTTGAACACAAAGCGCCAGAGATGATG 1112
QY 61 GGGAGGCGAGGGTGAAGTGGGCTCTGAGAGATGAGAGCTGCCAACCTTTGGCCAA 120
Db 1113 GGGAGGCGAGGGTGAAGTGGGCTCTGAGAGATGAGAGCTGCCAACCTTTGGCCAA 1172
QY 121 GCCCGCTCATGATCAAAAGCTCTAAGAGAAACAGAGCTGGCCCTTGTCCCTGACGGCGGACC 180
Db 1173 GCCCGCTCATGATCAAAAGCTCTAAGAGAAACAGAGCTGGCCCTTGTCCCTGACGGCGGACC 1232
QY 181 AGATGGTCAGTGCCTTGTGGATGCTGAGCCCCCATACTCTATTCCGAGTATGATCCTA 240
Db 1233 AGATGGTCAGTGCCTTGTGGATGCTGAGCCCCCATACTCTATTCCGAGTATGATCCTA 1292
QY 241 CCAGACCTTCAGTGAAGCTTCGATGATGGGCTTACTGACCAACCTGGCAGACGGGAGC 300
Db 1293 CCAGACCTTCAGTGAAGCTTCGATGATGGGCTTACTGACCAACCTGGCAGACGGGAGC 1352

RESULT 8
LOCUS AX474709 2092 bp DNA linear PAT 12-AUG-2002
DEFINITION Sequence 9 from Patent WO222882.
ACCESSION AX474709
VERSION AX474709.1 GI:22214043
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Stanton,M., Epstein,D. and Hamaguchi,N.
TITLE Target activated nucleic acid biosensor and methods of using same

JOURNAL Patent: WO 022882-A 9 21-MAR-2002;
Archemix Corporation (US)
FEATURES Location/Qualifiers
source 1..2092
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN

Query Match 100.0%; Score 336; DB 6; Length 2092;
Best Local Similarity 100.0%; Pred. No. 2.1e-85;
Matches 336; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GGATACGAAAGACCGAAGAGAGGAGGAGATGTTGAACACAAAGCGCCAGAGATGATG 60
Db 1053 GGATACGAAAGACCGAAGAGAGGAGGAGATGTTGAACACAAAGCGCCAGAGATGATG 1112
QY 61 GGGAGGCGAGGGTGAAGTGGGCTCTGAGAGATGAGAGCTGCCAACCTTTGGCCAA 120
Db 1113 GGGAGGCGAGGGTGAAGTGGGCTCTGAGAGATGAGAGCTGCCAACCTTTGGCCAA 1172
QY 121 GCCCGCTCATGATCAAAAGCTCTAAGAGAAACAGAGCTGGCCCTTGTCCCTGACGGCGGACC 180
Db 1173 GCCCGCTCATGATCAAAAGCTCTAAGAGAAACAGAGCTGGCCCTTGTCCCTGACGGCGGACC 1232
QY 181 AGATGGTCAGTGCCTTGTGGATGCTGAGCCCCCATACTCTATTCCGAGTATGATCCTA 240
Db 1233 AGATGGTCAGTGCCTTGTGGATGCTGAGCCCCCATACTCTATTCCGAGTATGATCCTA 1292
QY 241 CCAGACCTTCAGTGAAGCTTCGATGATGGGCTTACTGACCAACCTGGCAGACGGGAGC 300
Db 1293 CCAGACCTTCAGTGAAGCTTCGATGATGGGCTTACTGACCAACCTGGCAGACGGGAGC 1352
QY 301 TGGTTTCATGATCAACTGGGCGAAGAGGGTGCCAG 336
Db 1353 TGGTTTCATGATCAACTGGGCGAAGAGGGTGCCAG 1388

RESULT 9

LOCUS HUMERMCF 2092 bp mRNA linear PRI 08-NOV-1994
DEFINITION Human estrogen receptor mRNA, complete cds.
ACCESSION M12674
VERSION M12674.1 GI:182192
KEYWORDS estrogen receptor.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2092)
AUTHORS Greene,G.L., Gilna,P., Waterfield,M., Baker,A., Hort,Y. and Shine,J.
TITLE Sequence and expression of human estrogen receptor complementary DNA
JOURNAL Science 231 (4742), 1150-1154 (1986)
MEDLINE 86122927
PUBMED 3753802
COMMENT Original source text: Human breast cancer cell line MCF-7, cDNA to mRNA, clone OR8.
FEATURES Location/Qualifiers
source 1..2092
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/map="6q24-q27"
/cell_line="MCF-7"
/lab_host="Prokaryotae"
/tissue_lib="lambda gt11"
gene 1..2092
/gene="ESR"
mRNA <1..>2092
/gene="ESR"


```
RESULT 13
LOCUS      BD187708          2532 bp      DNA      linear      PAT 17-JUL-2003
DEFINITION Specific cleavage method of fusion protein.
ACCESSION  BD187708
VERSION    BD187708.1  GI:32997447
KEYWORDS   JP 2003024095-A/1.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE  1 (bases 1 to 2532)
AUTHORS    Nishiya, Y. and Oka, M.
TITLE      Specific cleavage method of fusion protein
JOURNAL    Patent: JP 2003024095-A 1 28-JAN-2003;
            TOYOBO CO LTD
COMMENT     OS Homo sapiens (human)
            PN JP 2003024095-A/1
            PD 28-JAN-2003
            PF 10-JUL-2001 JP 2001209476
            PI YOSHIAKI NISHIYA, MASANORI OKA
            PC C12P21/06//C12N15/09,C12N15/00
            CC Specific cleavage method of fusion protein
            FH Key
            FT source
            FT 1..2532
               Location/Qualifiers
               /organism="Homo sapiens (human)".
               /mol_type="genomic DNA"
               /db_xref="taxon:9606"

FEATURES             source
source              1..2532
                   /organism="Homo sapiens"
                   /mol_type="genomic DNA"
                   /db_xref="taxon:9606"

ORIGIN
Query Match      100.0%; Score 336; DB 6; Length 2532;
Best Local Similarity 100.0%; Pred. No. 2.1e-85;
Matches 336; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGATACGAAAGACCGAAGAGGAGGAGGAGATGTTGAAACACACAGCGCCGAGAGATGATG 60
Db 1505 GGATACGAAAGACCGAAGAGGAGGAGGAGATGTTGAAACACACAGCGCCGAGAGATGATG 1564
QY 61 GGGAGGCGAGGGTGGAAGTGGGTCTGCTGGAGACATGAGAGTGCACACCTTTGGCCAA 120
Db 1565 GGGAGGCGAGGGTGGAAGTGGGTCTGCTGGAGACATGAGAGTGCACACCTTTGGCCAA 1624
QY 121 GCCCGCTCATGATCAAAACGCTCTAAGAGAACAGCGCTTGTCCTTGACGGCCGACC 180
Db 1625 GCCCGCTCATGATCAAAACGCTCTAAGAGAACAGCGCTTGTCCTTGACGGCCGACC 1684
QY 181 AGATGGTCAGTGCCTTGTTGGATGCTGAGCCCCCCTACTCTATTCCGAGTATGATCCTA 240
Db 1685 AGATGGTCAGTGCCTTGTTGGATGCTGAGCCCCCCTACTCTATTCCGAGTATGATCCTA 1744
QY 241 CCAGACCCCTTCAGTGAAGCTTGATGATGGCTTACTGACCAACCTGGCAGACAGGGAGC 300
Db 1745 CCAGACCCCTTCAGTGAAGCTTGATGATGGCTTACTGACCAACCTGGCAGACAGGGAGC 1804
QY 301 TGGTTCACATGATCAACTGGGCGAGAGGGTGCCAG 336
Db 1805 TGGTTCACATGATCAACTGGGCGAGAGGGTGCCAG 1840

RESULT 14
LOCUS      HSM807087          5439 bp      mRNA      linear      PRI 28-AUG-2003
DEFINITION Homo sapiens mRNA; cDNA DKFZp686N23123 (from clone DKFZp686N23123);
            complete cds.
ACCESSION  BX640939
VERSION    BX640939.1  GI:34365270
KEYWORDS   Homo sapiens (human)
SOURCE     Homo sapiens
ORGANISM   Homo sapiens
```

```
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE  1 (bases 1 to 5439)
AUTHORS    Lauber, J., Bahr, A., Mewes, H.W., Weil, B., Amid, C., Osanger, A.,
            Fobo, G., Han, M. and Wiemann, S.
TITLE      The German Human cDNA Consortium
JOURNAL    Direct Submission
            Submitted (26-AUG-2003) MIPS, Ingolstaedter Landstr.1, D-85764
            Neuherberg, GERMANY
COMMENT     Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
            sequenced by Qiagen (Hilden/Germany) within the cDNA sequencing
            consortium of the German Genome Project.
            This clone (DKFZp686N23123) is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
            information about the clone and the sequencing project is available
            at http://mips.gsf.de/proj/cDNA/.

FEATURES             Location/Qualifiers
source              1..5439
                   /organism="Homo sapiens"
                   /mol_type="mRNA"
                   /db_xref="taxon:9606"
                   /map="6q25.1"
                   /clone="DKFZp686N23123"
                   /tissue_type="human uterus"
                   /clone_lib="686 (synonym: hlcc3). Vector pSport1_sfi; host
                   DH10B; sites SfiIA + SfiIB"
                   /dev_stage="adult"
                   1..5439
                   /gene="DKFZp686N23123"
                   234..1166
                   /gene="DKFZp686N23123"
                   /note="estrogen receptor alpha, differentially spliced"
                   /codon_start=1
                   /product="hypothetical protein"
                   /protein_id="CAE45969.1"
                   /db_xref="GI:34365271"
                   /translation="NAMESAKETRYCAVNCNDYASGVHYGVSCGCKAFKRSIQHN
                   DYMCATNCTIDKRRKCKYCEVGMKGGIRKRRGRLMKHKRQDDGE
                   GRPEGSAGDMRAANLPSPMLKSKNSLALSTADQWVHLDAEPPILYSEYDP
                   TRPFEASMMGLLTNLADRELRLMINWAKRVPFVDLTLDQVHLLCECHWLEITGL
                   VNRSEHPKLLFAPNLLDRNQGCKVEGMVEIFDMLLATTSRFRMNLQGBFVCLK
                   STILLNSGISHVEAKRILNLPKIFGNKWFPRV"
                   5416
                   /gene="DKFZp686N23123"

polyA_site
ORIGIN
Query Match      100.0%; Score 336; DB 9; Length 5439;
Best Local Similarity 100.0%; Pred. No. 2.1e-85;
Matches 336; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGATACGAAAGACCGAAGAGGAGGAGGAGATGTTGAAACACACAGCGCCGAGAGATGATG 60
Db 475 GGATACGAAAGACCGAAGAGGAGGAGGAGATGTTGAAACACACAGCGCCGAGAGATGATG 534
QY 61 GGGAGGCGAGGGTGGAAGTGGGTCTGCTGGAGACATGAGAGTGCACACCTTTGGCCAA 120
Db 535 GGGAGGCGAGGGTGGAAGTGGGTCTGCTGGAGACATGAGAGTGCACACCTTTGGCCAA 594
QY 121 GCCCGCTCATGATCAAAACGCTCTAAGAGAACAGCGCTTGTCCTTGACGGCCGACC 180
Db 595 GCCCGCTCATGATCAAAACGCTCTAAGAGAACAGCGCTTGTCCTTGACGGCCGACC 654
QY 181 AGATGGTCAGTGCCTTGTTGGATGCTGAGCCCCCCTACTCTATTCCGAGTATGATCCTA 240
Db 655 AGATGGTCAGTGCCTTGTTGGATGCTGAGCCCCCCTACTCTATTCCGAGTATGATCCTA 714
QY 241 CCAGACCCCTTCAGTGAAGCTTGATGATGGCTTACTGACCAACCTGGCAGACAGGGAGC 300
Db 715 CCAGACCCCTTCAGTGAAGCTTGATGATGGCTTACTGACCAACCTGGCAGACAGGGAGC 774
QY 301 TGGTTCACATGATCAACTGGGCGAGAGGGTGCCAG 336
```

Db 775 TGGTTCACATGATCACTGGCGGAAGAGGTGCCAG 810

RESULT 15
AR153585
LOCUS AR153585 6450 bp DNA linear PAT 08-AUG-2001
DEFINITION Sequence 34 from patent US 6235872.
ACCESSION AR153585
VERSION AR153585.1 GI:15121117
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.

REFERENCE 1 (bases 1 to 6450)
AUTHORS Bredesen,D.E. and Razizadeh,S.
TITLE Proapoptotic peptides dependence polypeptides and methods of use
JOURNAL Patent: US 6235872-A 34 22-MAY-2001;
FEATURES Location/Qualifiers
source 1..6450
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN
Query Match 100.0%; Score 336; DB 6; Length 6450;
Best Local Similarity 100.0%; Pred. No. 2.1e-85;
Matches 336; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GGATACGAAAGACCGAAGAGAGGGAGATGTTGAAACACAAGCGCCAGAGAGATGATG 60
Db 1121 GGATACGAAAGACCGAAGAGAGGGAGATGTTGAAACACAAGCGCCAGAGAGATGATG 1180
QY 61 GGGAGGGCAGGGGTGAAGTGGGGTCTGCTGGAGACATGAGAGCTGCCAACCTTTGGCCAA 120
Db 1181 GGGAGGGCAGGGGTGAAGTGGGGTCTGCTGGAGACATGAGAGCTGCCAACCTTTGGCCAA 1240
QY 121 GCCCGCTCATGATCAACAGCTCTAAGAAAGACAGAGCTGGCCTTGCCCTGACGGCCGACC 180
Db 1241 GCCCGCTCATGATCAACAGCTCTAAGAAAGACAGAGCTGGCCTTGCCCTGACGGCCGACC 1300
QY 181 AGATGGTCAGTCCCTTGTTGGATGCTGAGCCCCCCTACTCTATTCCGAGTATGATCCTA 240
Db 1301 AGATGGTCAGTCCCTTGTTGGATGCTGAGCCCCCCTACTCTATTCCGAGTATGATCCTA 1360
QY 241 CCAGACCCCTTCAGTGAAGCTTCGATGATGGGCTTACTGACCAACCTGGCAGACAGGGAGC 300
Db 1361 CCAGACCCCTTCAGTGAAGCTTCGATGATGGGCTTACTGACCAACCTGGCAGACAGGGAGC 1420
QY 301 TGGTTCACATGATCACTGGCGGAAGAGGGTGCCAG 336
Db 1421 TGGTTCACATGATCACTGGCGGAAGAGGGTGCCAG 1456

Search completed: November 3, 2004, 07:43:08
Job time : 2097.38 secs